

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Informática

Departamento de Ingeniería del Software e Inteligencia Artificial (DISIA)



Proyecto Fin de Máster en Sistemas Inteligentes

Máster en Investigación en Informática

Determinación taxonómica inteligente de Flora Ibérica mediante lógica Fuzzy.

Diana Díaz Agrela

Director:

Dr. D. Luis Garmendia Salvador

Codirector:

Dr. D. Alfonso Garmendia Salvador

Universidad Politécnica de Valencia, España.

Instituto Agroforestal Mediterráneo

Madrid (España), 2009

La abajo firmante, matriculada en el Máster en Investigación en Informática de la Facultad de Informática, autoriza a la Universidad Complutense de Madrid (UCM) a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a su autor el presente Trabajo Fin de Máster: "Determinación taxonómica inteligente de Flora Ibérica mediante lógica Fuzzy", realizado durante el curso académico 2008-2009 bajo la dirección de Dr. D. Luis Garmendia Salvador y con la colaboración externa de dirección de Dr. D. Alfonso Garmendia Salvador en el Departamento de Ingeniería del Software e Inteligencia Artificial (DISIA), y a la Biblioteca de la UCM a depositarlo en el Archivo Institucional E-Prints Complutense con el objeto de incrementar la difusión, uso e impacto del trabajo en Internet y garantizar su preservación y acceso a largo plazo.

Madrid, Septiembre 2009

Fmdo: Diana Díaz Agrela

Resumen

Ante la dificultad que existe para la correcta determinación de un ejemplar concreto de cualquiera de las múltiples especies de plantas existentes, se ha investigado la utilización de un instrumento novedoso, la teoría de conjuntos borrosos y el razonamiento aproximado. La información de las características de los distintos taxones se representa mediante conjuntos borrosos. Dichos atributos son las entradas de un sistema inteligente de inferencia, siendo las salidas un grado de pertenencia para cada posible taxón al que puede pertenecer el ejemplar a determinar, incluso en casos con información incompleta o falta de datos requeridos.

El estudio previo pretende encontrar una representación del conocimiento recogido en las claves y descripciones de Flora Ibérica. A partir de las características y taxones, se procede a un análisis de la información y su posterior estructuración en categorías de características, detección de asociaciones entre ellas y evaluación numérica de sus relaciones con los taxones.

Previamente al diseño del sistema borroso, se implementa un sistema nítido que establece los requisitos para cualquier sistema genérico de determinación taxonómica. Un requisito destacable es dotar al sistema de flexibilidad, de tal manera que su conocimiento es auto-actualizable, las modificaciones surten efecto inmediato en la interface gráfica presentada. Así mismo se calculan nuevos conjuntos borrosos de forma automática y se modifican las reglas de inferencia de cada taxón.

Abstract

Taxonomic determination is an area open for improvement by research because of its difficulty, fuzzy logic and approximate reasoning are appropriate and innovative tools. The present study used fuzzy logic to implement an intelligent system for identifying specimens of *Flora Ibérica*. Fuzzy sets represent various taxonomic characteristics, which are inputs of an intelligent inference engine using fuzzy rules of inference to determine the specimen, even in cases with absent or incomplete required information.

The prior study seeks to find a representation of knowledge contained in the keys and descriptions of *Flora Ibérica*. From the characteristics and taxa, we proceed to an information analysis and its subsequent structure into categories of characteristics, detection of associations among them and numerical evaluation of its relations with the taxa.

Prior to the fuzzy system design, we implement a clear system that fixes requirements for any generic system of taxonomic determination. A remarkable requirement is to design a flexible system with knowledge self-updating, then any change takes effect immediately in the graphical interface submitted. Also, new fuzzy sets are calculated automatically and the inference rules of each taxon are updated.

Índice general

1. Introducción	1
2. Estado del arte	5
2.1. Descripciones de flora y claves	5
2.2. Clave dicotómica	7
2.3. Algunos sistemas de determinación	9
2.4. Estudio previo. Prímulas	11
3. Material y métodos	13
3.1. Lógica borrosa	13
3.1.1. Conjuntos borrosos	13
3.1.2. Lógica borrosa	14
3.1.3. Etiquetas lingüísticas	14
3.1.4. Inferencia y razonamiento aproximado	15
3.2. XFUZZY	15
4. Determinación taxonómica. Conocimiento	17
4.1. Conocimiento. Flora (Taxones).	19
4.2. Conocimiento. Características.	20
4.2.1. Agrupaciones de características	20
4.2.2. Características Cuantitativas	21
4.2.3. Características Cualitativas	23

4.3.	Representación numérica del lenguaje científico	28
4.4.	Representación del conocimiento	28
4.4.1.	Sistema adaptativo	28
4.4.2.	Introducción de datos	29
4.4.3.	Traducción a metalenguaje estructurado, XML	37
5.	<i>Sistema inteligente nítido de determinación taxonómica</i>	39
5.1.	Selección de respuestas	39
5.2.	Inferencia nítida	44
5.2.1.	Premisas a evaluar y valor a asignar a dicha evaluación	44
5.2.2.	Ejecución de reglas	48
5.2.3.	Algoritmos	49
6.	<i>Sistema inteligente borroso de determinación taxonómica</i>	51
6.1.	Características	52
6.1.1.	Características Cuantitativas	52
6.1.2.	Características Cualitativas	60
6.2.	Inferencia borrosa	74
6.2.1.	Fuzzy set inferido	74
6.2.2.	Fuzzy set fundamentado	75
6.2.3.	Toma de decisiones	76
7.	Conclusiones	77
8.	Trabajo Futuro	81
9.	Bibliografía	85
10.	Apéndice 1. Introducción a la lógica borrosa	89
11.	Apéndice 2. Tabla Excel de conocimiento de determinación taxonómica.	103
12.	Apéndice 3. Descripción de esquemas XML. Representación del conocimiento	113

Índice de figuras

<i>Figura 2-1. Ejemplo de ilustración para la familia Compositae (Asteraceae) [dicotiledóneas].</i>	6
<i>Figura 2-2. Familia, SOLANACEAE. Género : Datura. Imágenes de izquierda a derecha, flor, fruto, imagen de campo.</i>	7
<i>Figura 2-3. Clave General. Flora Ibérica. Ejemplo.</i>	8
<i>Figura 2-4. Ejemplo de claves de la pteridoflora de las Islas Canarias.</i>	8
<i>Figura 2-5. Ejemplo de ilustración. D, extremo de una rama de Selaginella denticulata con microsporangios y megasporangios</i>	9
<i>Figura 2-6. Ejemplo de uso de la herramienta online Los macrolíquenes epífitos de España. Una guía interactiva.</i>	9
<i>Figura 2-7. Herbario virtual de las Islas Baleares.</i>	10
<i>Figura 2-8 Pantalla con la interface gráfica del programa de identificación de Prímulas.</i>	12
<i>Figura 3-1. Ejemplo de modelo de variables lingüísticas</i>	15
<i>Figura 3-2. Flujo de Diseño de Xfuzzy 3.0.</i>	16
<i>Figura 4-1. Esquema ilustrativo general de los elementos involucrados en la determinación.</i>	18
<i>Figura 4-2. Proceso genérico de sistemas de determinación.</i>	18
<i>Figura 4-3. Esquema jerárquico de la estructuración del conocimiento de Flora Ibérica.</i>	19
<i>Figura 4-4. Gráfica del conjunto de límites (Min, MinInt, MaxInt, Max) para una relación $R_{xx}(T_x, C_x)$.</i>	22
<i>Figura 4-5. Gráfica del conjunto de límites (Med, ES) para una relación $R_{xx}(T_x, C_x)$.</i>	22
<i>Figura 4-6. Tipos de taxón.</i>	30
<i>Figura 4-7. Jerarquía representada en el ejemplo de la Características</i>	31

<i>Figura 5-1. Ejemplo de selección de respuesta a una característica cualitativa.</i>	40
<i>Figura 5-2. Ejemplo de selección de características cualitativas.</i>	41
<i>Figura 5-3. Ejemplo de selección sí, no, no sé de características cualitativas.</i>	42
<i>Figura 5-4. Ejemplo de un grupo de auto-exclusión, solo una posible selección.</i>	43
<i>Figura 5-5. Ejemplo de un grupo de auto-exclusión con selección sí, no, no sé.</i>	43
<i>Figura 5-6. Valores inferidos en intervalos delimitados por límites (Min, MinInt, MaxInt, Max) para una relación $R_{xx}(T_x, C_x)$. Tres ejemplos (en este orden) con todos los límites, sin un límite, solo con un límite.</i>	45
<i>Figura 5-7. Valores inferidos en intervalos delimitados por límites (Med, ES).</i>	45
<i>Figura 5-8. Ejemplo de visualización en una posible implementación.</i>	50
<i>Figura 6-1. Función de pertenencia trapezoidal genérica.</i>	54
<i>Figura 6-2. Ejemplo conjunto borroso mf_Abies de la característica t_Altura_m.</i>	55
<i>Figura 6-3. Ejemplo conjunto borroso mf_Abies de la característica t_Longitud_estróbilos_cm.</i>	56
<i>Figura 6-4. Ejemplo conjunto borroso mf_PINACEAE de la característica t_Longitud_estróbilos_cm.</i>	56
<i>Figura 6-5. Intervalo máximo. Ejemplo conjunto borroso mf_Abies de la característica t_Longitud_de_las_hojas_mm.</i>	57
<i>Figura 6-6. Intervalo máximo. Ejemplo conjunto borroso mf_Pinus_halepensis de la característica t_Longitud_de_las_hojas_mm.</i>	57
<i>Figura 6-7. Función de pertenencia trapezoidal genérica.</i>	59
<i>Figura 6-8. Intervalo frecuente. Ejemplo conjunto borroso mf_Pinus_halepensis de la característica t_Longitud_de_las_hojas_mm.</i>	59
<i>Figura 6-9. Ejemplo conjuntos borrosos mf_no y mf_si de la característica booleana t_Árbol.</i>	60
<i>Figura 6-10. Ejemplo conjuntos borrosos mf_no_soft de la característica booleana t_Árbol.</i>	63
<i>Figura 6-11. Ejemplo conjuntos borrosos mf_si_soft de la característica booleana t_Árbol.</i>	63
<i>Figura 6-12. Posible implementación para características aproximadas.</i>	64
<i>Figura 6-13. Ejemplo conjuntos borrosos en características aproximadas.</i>	65
<i>Figura 6-14. Posible implementación de características auto-excluyentes relacionadas aproximadas.</i>	69
<i>Figura 6-15. Conjuntos borrosos trapezoidales clásicos.</i>	70

<i>Figura 6-16. Esquema de características relacionadas aproximadas, y sus áreas de mayor y menor grado de “se da” dichas características.</i>	72
<i>Figura 6-17. Conjuntos borrosos de tres características relacionadas aproximadas.</i>	73
<i>Figura 6-18. Conjuntos borrosos “Extremadamente poco de” para las características Copa redondeada y Copa cónica o piramidal en este orden.</i>	73
<i>Figura 6-19. Función de pertenencia del conjunto borroso fundamentado.</i>	75
<i>Figura 12-1. Esquemas XML</i>	113
<i>Figura 12-2. Conocimiento_caracteristicas</i>	114
<i>Figura 12-3. Conocimiento_flora</i>	121

Índice de tablas

<i>Tabla 4-1. Descripción de límites de características cuantitativas.</i>	22
<i>Tabla 4-2. Posibilidades positivas asociadas a una relación $R_{xx}(T_x, C_x)$.</i>	24
<i>Tabla 4-3. Conjunto completo de posibilidades asociadas a una relación $R_{xx}(T_x, C_x)$.</i>	24
<i>Tabla 4-4 Ejemplo de estructuración de taxones.</i>	30
<i>Tabla 4-5. Reglas a utilizar para establecer la relación jerárquica de taxones.</i>	31
<i>Tabla 4-6. Ejemplo de estructuración de agrupaciones.</i>	31
<i>Tabla 4-7. Ejemplo de agrupaciones ocultas.</i>	32
<i>Tabla 4-8. Representación gráfica de las relaciones para características cualitativas.</i>	33
<i>Tabla 4-9. Representación gráfica de las relaciones para características cuantitativas.</i>	33
<i>Tabla 4-10. Ejemplo de características.</i>	34
<i>Tabla 4-11. Ejemplo de características cualitativas booleanas y aproximadas.</i>	35
<i>Tabla 4-12. Ejemplo características auto-excluyentes.</i>	36
<i>Tabla 4-13. Ejemplo de características auto-excluyentes relacionadas aproximadas</i>	37
<i>Tabla 5-1. Características no unificadas.</i>	41
<i>Tabla 5-2. Conjunto completo de posibilidades asociadas a una relación $R_{xx}(T_x, C_x)$ para una característica cualitativa.</i>	44
<i>Tabla 5-3. Ejemplo baremos características auto-excluyentes con respuestas negativas.</i>	47
<i>Tabla 5-4. Ejemplo conjunto de características seleccionadas.</i>	48
<i>Tabla 5-5. Matriz de valores para las relaciones.</i>	48

<i>Tabla 5-6. Ejemplo conjunto de características con respuestas sí, no, ni sé.</i>	49
<i>Tabla 6-1. Representación gráfica de las relaciones para características cuantitativas.</i>	52
<i>Tabla 6-2. Ejemplo 1. Altura (m)</i>	54
<i>Tabla 6-3. Puntos que describen el trapecio de cada una de las funciones de pertenencia de los conjuntos borrosos de los taxones Abies y Pinus.</i>	54
<i>Tabla 6-4. Ejemplo 2. Longitud de los estróbilos (cm)</i>	55
<i>Tabla 6-5. Puntos que describen el trapecio de cada una de las funciones de pertenencia de los conjuntos borrosos de los taxones PINACEAE y Abies.</i>	55
<i>Tabla 6-6. Ejemplo 3. Longitud de las hojas (mm)</i>	56
<i>Tabla 6-7. Puntos que describen el trapecio de cada una de las funciones de pertenencia de los conjuntos borrosos de los taxones Abies y Pinus halapensis.</i>	57
<i>Tabla 6-8. Descripción de límites de características cuantitativas.</i>	58
<i>Tabla 6-9. Intervalo frecuente. Puntos que describen el trapecio de cada una de las funciones de pertenencia de los conjuntos borrosos de los taxones Abies y Pinus halapensis.</i>	59
<i>Tabla 6-10. Ejemplo de característica booleana</i>	61
<i>Tabla 6-11. Conjunto completo de posibilidades asociadas a una relación $R_{ij}(T_i, C_j)$.</i>	66
<i>Tabla 6-12. Ejemplo de característica aproximada, fusiformes.</i>	66
<i>Tabla 6-13. Ejemplo de grupo relacionado aproximado de 2 características.</i>	70

Capítulo 1

Introducción

La determinación taxonómica de especímenes es una tarea que siempre ha estado presente en el trabajo de los botánicos. En los últimos tiempos se han desarrollado nuevas disciplinas que necesitan de esta tarea, por ejemplo en la realización de estudios de impacto ambiental cobra especial importancia el inventariado taxonómico de las especies encontradas. Para un botánico, sea cual sea su rama, es muy importante poder realizar determinaciones lo más acertadas posibles, para lo cual se suele recurrir a las claves dicotómicas y a las descripciones técnicas de las especies recogidas en las "floras". En cualquiera de estos casos, es imprescindible poder reconocer las diferentes especies con una eficacia suficiente. Esta tarea puede ser muy dificultosa dado que, aquello que puede resultar relativamente fácil para los profesionales que llevan tiempo trabajando en una determinada área, resulta complicado al cambiar de zona. La dificultad aumenta si para la determinación no se cuenta con un botánico cerca al que realizar la consulta.

La determinación está basada generalmente en los caracteres o atributos morfológicos, que pueden tener tipos muy variados, conteniendo muchos de ellos elementos con incertidumbre o falta de precisión. Las características que el agente determinista tiene que observar, pueden no aparecer claramente a los ojos del observador, lo que puede confundirle y provocar que no sea capaz de emitir un juicio consistente. Tanto las definiciones de las características a observar, como las posibles respuestas que puede inferir el observador, están definidas con lenguaje natural en ocasiones poco preciso. Por lo tanto, el razonamiento utilizado para la inferencia de la solución está basado en

unos elementos de partida que contienen cierta incertidumbre, tanto en su definición como en el modo de ser observados. Así mismo, las respuestas emitidas sufren también de indeterminación. Podemos percibir este razonamiento como razonamiento aproximado. Pensemos, por ejemplo, si la clave separa entre hojas lisas o rugosas, ¿cómo de rugosa tiene que ser la hoja para incluirla en la segunda categoría?.

Los métodos Dicotómicos clásicos emiten una serie de preguntas con únicamente dos posibles respuestas sí o no, cada respuesta enlaza seguidamente con otra nueva pregunta. Nótese que no es posible eludir ninguna de las preguntas del orden preestablecido. En el caso de que el observador no sea capaz de evaluar la pregunta no podrá continuar con la determinación. Estos métodos provocan que sea necesaria bastante experiencia para ser capaz de utilizar las claves con la soltura necesaria, ya que el botánico, sin ser consciente de ello, utiliza técnicas del sentido común o del razonamiento aproximado para tomar sus decisiones.

La lógica borrosa es capaz de modelar y utilizar el conocimiento incompleto, como el que encontramos en el ámbito de la determinación taxonómica. Mediante conjuntos borrosos podemos diseñar funciones de pertenencia para cada característica, que modelen las respuestas imprecisas del lenguaje natural como poco, medio, mucho. Así mismo, las relaciones excluyentes entre algunas características y los juicios aproximados pueden convertirse en elementos del sistema, que permitan extraer información adicional sobre la propia característica.

El sistema diseñado intenta paliar otro problema típico de los métodos clásicos dicotómicos. Acabamos de comentar, el modo en que estos métodos tienen un orden preestablecido con varios caminos posibles descritos en todo momento por la respuesta negativa o positiva que emita el usuario. En consecuencia, el usuario nunca puede eludir una pregunta si la desconoce. Por ejemplo, si en la clave se pregunta por el número de pétalos o por su color, y el ejemplar que se quiere determinar no tiene flores, es imposible continuar y hay que desistir de la determinación del espécimen o recurrir a la lectura de todas las descripciones en las floras, para todas las especies posibles, lo cual se hace imposible si el conjunto de posibilidades es grande. Adicionalmente, si la persona que desea determinar la especie no es un experto, puede no comprender la pregunta y por tanto desconocer la respuesta, como por ejemplo si se pregunta si la base está bruscamente contraída y no conoce el significado de ser “bruscamente contraída”. Para solventar esta problemática el sistema debe permitir al usuario introducir, en el orden deseado por él mismo, las respuestas a las características evaluadas; debe también proporcionar herramientas que le permitan rectificar las respuestas a preguntas ya contestadas o incluso eliminarlas de forma independiente al resto.

La primera dificultad encontrada en la realización de este sistema ha sido la representación del conocimiento de la Flora, como hemos visto, los expertos suelen hacer uso de claves y de descripciones. El objetivo es modelar este conocimiento para que pueda convertirse en la entrada de un sistema de determinación. En este estudio se ha llevado a cabo una normalización de las características encontradas tal que, partiendo de un lenguaje científico (botánico) expresado en lenguaje natural, obtenemos una representación numérica y estructurada del mismo. De esta manera, conseguimos extraer el máximo grado de información tanto de la definición propia de cada característica, como de sus relaciones con otras características y sus posibles valores observables.

Esta normalización del conocimiento taxonómico se ha diseñado con gran flexibilidad, imponiendo sobre el sistema el requisito de ser un sistema adaptativo. El conocimiento vendrá representado en tablas Excel estructuradas donde el experto será capaz de introducir nuevo conocimiento, modificar el previamente introducido y/o eliminar o corregir errores. Con el Empleo de tablas Excel se pretende facilitar el uso a los expertos que están frecuentemente familiarizados con este tipo de representación. El sistema posteriormente transforma la información a XML, este metalenguaje nos permite un intercambio de información estándar, lo que le proporciona apertura hacia futuras evoluciones.

El sistema es adaptativo, permitiendo adoptar todas las modificaciones de forma automática, tanto en la representación gráfica, la categorización de atributos y taxones, las premisas a evaluar, los conjuntos borrosos diseñados (en el sistema borroso) y las reglas de inferencia para cada taxón. Por ejemplo, si el usuario introduce una nueva característica con todas sus consecuentes relaciones con los taxones, el sistema es capaz de representar la característica según el tipo que sea y sus relaciones con otras características; generar y evaluar las premisas en las que se puede ver involucrada dicha característica con los distintos taxones y finalmente utilizarla como un elemento más de inferencia. Así mismo si se introduce un nuevo taxón, se generará automáticamente sus reglas de inferencia y podrá formar parte del conjunto solución de la determinación.

Esta flexibilidad proporciona una inagotable alimentación de conocimiento, ya que diferentes expertos podrían incrementar su propia información y además mezclarla con la de otros expertos. Se podría alcanzar una base de información sólida y de gran completitud, disponible para cualquier usuario del sistema, por ejemplo vía Web.

La implementación del sistema “DetFlora” puede encontrarse en la Web [12] .

Capítulo 2

Estado del arte

2.1. Descripciones de flora y claves

Como se verá en este estudio, la normalización del conocimiento se basa en extraer información de la literatura conocida y en base a los datos recogidos, realizar una representación de su contenido que permita ser utilizada por un sistema inteligente.

Para este trabajo nos hemos centrado principalmente en el apartado *Gimnospermas* del Tomo I de Flora Ibérica [10]. Este proyecto se inició en 1980 y con sus propias palabras *“pretende – mediante investigación taxonómica original – poner al día y sintetizar los conocimientos actuales sobre las plantas vasculares que crecen espontáneamente en la Península Ibérica e Islas Baleares, territorio de una notable riqueza florística. El objetivo último es publicar una obra – tanto en forma de libro, como en formato electrónico, CD-ROM o incluso distribuido en Internet “*. Esta iniciativa alcanza en estos momentos el 70% de la Flora Ibérica, según comentan, aproximadamente en 8 o 10 años podrán concluir con éxito la tarea.

La principal ventaja que obtenemos al utilizar esta base de conocimiento es precisamente la calidad y precisión de los datos aportados. La actualización que han llevado a cabo afecta tanto a novedades taxonómicas como a nuevas nomenclaturas adoptadas. Realizar un trabajo en función del proyecto Flora Ibérica, nos proporciona confianza respecto a la completitud y adecuación de nuestra representación de dicho conocimiento. El uso de una nomenclatura actualizada, produce un acercamiento mayor a los expertos botánicos que puedan ser futuros usuarios del sistema, al utilizar un lenguaje conocido para ellos.

Una posible evolución del sistema podría ser incluir, en la medida de lo posible, imágenes que pudieran facilitar la identificación de determinadas características enunciadas al usuario. Los sistemas que incluyen herramientas gráficas flexibilizan su uso a usuarios menos expertos; en ocasiones proporcionando medios gráficos y descripciones literarias, un usuario menos experto puede servirse de estas herramientas para realizar una correcta determinación taxonómica.

El proyecto Flora Ibérica contiene información gráfica, destacamos otros proyectos que podrían ser de utilidad como la Asociación BIGA (Biodiversidad en Galicia) [13], dedicada al estudio del patrimonio natural de Galicia, se centrada sobre todo en trabajos sobre artrópodos y flora vascular. En su sitio Web dispone del espacio *Plantae* con información sobre flora vascular gallega, en la cual podemos encontrar una galería fotográfica por familia y género muy correcta, así como un conjunto de publicaciones online.

En el Herbario de la Universidad Pública de Navarra [15], podemos encontrar información de la flora arvense principalmente en el área de las malas hierbas, que es también una parte importante de la biodiversidad vegetal de nuestro territorio. El herbario contiene fichas detalladas por familia y una galería de imágenes muy interesante, ya que en algunas de las ilustraciones podemos encontrar esquemas de las partes informadas en la descripción. Podemos ver un ejemplo en la Figura 2-1.



Figura 2-1. Ejemplo de ilustración para la familia Compositae (Asteraceae) [dicotiledóneas].

El Herbario Virtual del Mediterráneo Occidental [16], recoge información y una amplia galería de imágenes de las plantas vasculares de las tierras de la cuenca del Mediterráneo Occidental. Encontramos especial interés en las imágenes, ya que para los distintos géneros proporcionan varias imágenes, como por ejemplo la flor, el futo y una imagen de campo. Podemos ver un ejemplo en la Figura 2-2.



Figura 2-2. Familia, SOLANACEAE. Género : *Datura*. Imágenes de izquierda a derecha, flor, fruto, imagen de campo.

El Proyecto Flora Ibérica “De lo real a lo imaginario”, Universidad de Valencia [17], contiene un buscador de flora por división, familia, género, especie y nombre. Otro proyecto muy interesante son las Lecciones Hipertextuales de Botánica [18], este proyecto se inició para facilitar el aprendizaje de la Botánica y recoge la información de una forma muy cómoda y estructurada.

2.2. Clave dicotómica

El método adicional de obtención de información son las claves dicotómicas, se han utilizado las comprendidas en Flora Ibérica. Las claves dicotómicas nos permiten extraer información más precisa y de gran relevancia en la determinación. Los métodos dicotómicos pretenden alcanzar la solución siempre por el camino más corto posible.

El problema de estos métodos es que el usuario no puede eludir ninguna pregunta. Se trata de un sistema que realiza preguntas con dos únicas repuestas, sí o no. Si el usuario no conoce la respuesta o bien el espécimen de estudio carece de las premisas para evaluar una pregunta (por ejemplo, preguntar por el color de los pétalos cuando el espécimen no tiene flores), la determinación tiene que finalizar sin éxito. En la Figura 2-3 puede verse un ejemplo de clave.

Angiospermae

Incluye todas las familias de los volúmenes I al VIII, más el X, XIII, XIV, XV, XVIII y XXI, que van numeradas, y la gran mayoría de las familias de los demás volúmenes, aunque puede haber sido omitida alguna de las tan solo representadas por plantas introducidas.

1.	Plantas acuáticas, sumergidas o flotantes, que no enraízan en el substrato	2
–	Plantas terrestres o acuáticas que enraízan en el substrato	7
2[1].	Plantas sin hojas ni tallos claramente diferenciados ... CLXXX. <i>Lemnaceae</i> (vol. 18)	
–	Plantas con hojas y tallos claramente diferenciados	3
3[2].	Hojas divididas en numerosos segmentos filiformes	4
–	Hojas no divididas en numerosos segmentos filiformes	5
4[3].	Plantas con pequeñas vejigas en las hojas, o tallos aparentemente afilos	
 CLII. <i>Lentibulariaceae</i> p.p. (vol. 14)	
–	Plantas sin pequeñas vejigas en las hojas; tallos siempre con hojas verticiladas o semiverticiladas	XXXIV. <i>Ceratophyllaceae</i> p.p. (vol. 1)
5[3].	Pecíolos no engrosados y con espacios intercelulares aeríferos	
 Hydrocharitaceae p.p. (vol. 17)	
–	Pecíolos, al menos los de las hojas emergentes, engrosados y con espacios intercelulares aeríferos.....	6

Figura 2-3. Clave General. Flora Ibérica. Ejemplo.

La Clave taxonómica de la pteridoflora de las Islas Canarias [14] encontramos un conjunto típico de claves dicotómicas para la determinación de los diferentes taxones de helechos de las Islas Canarias. Si nos fijamos en la Figura 2-4, el modo de representar las claves es similar al visto anteriormente.

CLAVE DE FAMILIAS

1.	Plantas que producen micrósporas y megásporas	2
1.	Plantas que producen un solo tipo de esporas	3
2.	Plantas subacuáticas, con esporangios reunidos en soros y éstos, a su vez, en esporocarpos	4
2.	Plantas no subacuáticas, de aspecto musgoso. Esporangios solitarios situados en las axilas de los esporófilos. Éstos se reúnen, a su vez, en espigas terminales.....	<i>Selaginellaceae</i>

Figura 2-4. Ejemplo de claves de la pteridoflora de las Islas Canarias.

También se incluyen algunas ilustraciones que representan aspectos morfológicos que se han tenido en cuenta en el trabajo. Estas ilustraciones podrían servirnos para ilustrar características complejas como hemos visto anteriormente, podemos ver un ejemplo en la Figura 2-5.

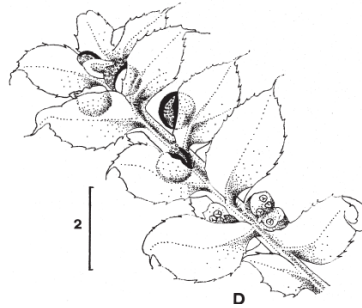


Figura 2-5. Ejemplo de ilustración. D, extremo de una rama de *Selaginella denticulata* con microsporangios y megasporangios

2.3. Algunos sistemas de determinación

El estudio de *Los macrolíquenes epífitos de España. Una guía interactiva* [19] , cuenta con una implementación online del método dicotómico clásico. Esta herramienta va realizando preguntas de dos en dos al usuario, el cual debe elegir la respuesta deseada para continuar con la elección. Como apoyo a la decisión, muestra en cada iteración imágenes de especímenes con las características descritas. Adicionalmente, muestra en cada paso al usuario cuantas especies quedan en el conjunto posible de soluciones. En todo momento nos da la opción de generar textualmente el conjunto completo de claves y descripciones para las especies posibles en la solución actual, en el ejemplo de la Figura 2-6, vemos como nos da la opción de ver las claves para 6 especies.



Figura 2-6. Ejemplo de uso de la herramienta online *Los macrolíquenes epífitos de España. Una guía interactiva*.

El Herbario virtual de las Islas Baleares [21], utiliza una herramienta de introducción de datos en la cual el usuario puede introducir las características observadas en el orden que desee. Así mismo observamos cómo hay detectadas dos tipos de características, aquellas con posibles respuestas sí y no, frente al resto en las que únicamente existe la respuesta sí. En nuestro estudio también se tienen en cuenta estos factores. Podemos observar también, como el aplicativo permite filtrar al usuario por nombre de familia o de especie si lo conoce. Una vez se rellenan los datos deseados, es necesario ejecutar el algoritmo para conseguir la determinación. Este proceso puede resultar pesado al usuario, ya que a menudo es deseable ver los “efectos” de nuestras elecciones inmediatamente. Podemos ver un ejemplo en la Figura 2-7. Herbario virtual de las Islas Baleares. Las diferentes características están dispuestas en agrupaciones superiores.

Herbari virtual de les Illes Balears
Laboratori de Botànica | Departament de Biologia

CLAU DE CLASSIFICACIÓ (PROTOTIP, V 2.1)

Seleccioneu les característiques de la planta a classificar i premeu el botó “Executar Cerca”, per a començar de nou premeu “Reset Cerca”

Executar Cerca Reset Cerca

Cerca per nom d'espècie

Cerca per nom de família

TIPUS DE PLANTA

Planta llenyosa ☐ Sí ☐ No

Planta herbàcia ☐ Sí ☐ No

Liana ☐ Sí ☐ No

MES DE FLORACIÓ

Desembre ☐ Sí ☐ No

Gener ☐ Sí ☐ No

Febrer ☐ Sí ☐ No

Març ☐ Sí ☐ No

Abril ☐ Sí ☐ No

Maig ☐ Sí ☐ No

Juny ☐ Sí ☐ No

Juliol ☐ Sí ☐ No

Agost ☐ Sí ☐ No

Setembre ☐ Sí ☐ No

Octubre ☐ Sí ☐ No

Novembre ☐ Sí ☐ No

HABITAT

1. Aiguamolls i salobrars ☐ Sí ☐ No

2. Camps de conreu, Vores de camins, llocs alterats ☐ Sí ☐ No

3. Costa rocosa ☐ Sí ☐ No

4. Llocs humits no salins (Ramblas, basses, torrents, canals, fonts i séquies) ☐ Sí ☐ No

5. Plantas de jardí ☐ Sí ☐ No

6. Pinars i garrigues amb xiprer i timonedes ☐ Sí ☐ No

7. Marines de bruc, brolles silícioles ☐ Sí ☐ No

8. Pradells terofítics calcícoles ☐ Sí ☐ No

9. Pradells terofítics silícioles ☐ Sí ☐ No

10. Platges i dunes, sòls arenosos ☐ Sí ☐ No

11. Alzinars ☐ Sí ☐ No

12. Brolles xero-sclerofítics de muntanya i carrixerers ☐ Sí ☐ No

13. Ullastrars i altres garrigues esclerofíles, sabinars ☐ Sí ☐ No

14. Esclerites de roques i penya-segats ☐ Sí ☐ No

15. Petits talussos ombrívols ☐ Sí ☐ No

16. Bardisses ☐ Sí ☐ No

17. Cultivades ☐ Sí ☐ No

18. Superfícies rocoses, parets seques i marges ☐ Sí ☐ No

COLOR DE LA FLOR

Flor Groc a crema ☐ Sí ☐ No

Flor Blanca ☐ Sí ☐ No

Flor Rosa a morat ☐ Sí ☐ No

Flor blava ☐ Sí ☐ No

Flor vermella ☐ Sí ☐ No

Flor verda ☐ Sí ☐ No

Flor marrón ☐ Sí ☐ No

TIPUS DE FLOR

Flor regular ☐ Sí ☐ No

Flor irregular ☐ Sí ☐ No

Flor reduïda ☐ Sí ☐ No

Flor hermafrodita ☐ Sí ☐ No

Flor unisexual ☐ Sí ☐ No

FORMA I MIDA DELS PÈTALS

Pètals evidents ☐ Sí ☐ No

Sense pètals evidents ☐ Sí ☐ No

Pètals: sencers ☐ Sí ☐ No

Pètals escotats ☐ Sí ☐ No

Pètals dentats ☐ Sí ☐ No

Pètals Lobulats o laciniats ☐ Sí ☐ No

Mida pètals: 1 a 2mm ☐ Sí ☐ No

de 2 a 5mm ☐ Sí ☐ No

de 5 a 10mm ☐ Sí ☐ No

de 10 a 20mm ☐ Sí ☐ No

Més de 20mm ☐ Sí ☐ No

Pètals_Tépals Lliures ☐ Sí ☐ No

Pètals_Tépals soldats ☐ Sí ☐ No

Sépals_Tépals lliures ☐ Sí ☐ No

Sépals_Tépals soldats ☐ Sí ☐ No

Figura 2-7. Herbario virtual de las Islas Baleares.

En *Biology 211: Taxonomy of Flowering Plants. World Wide Flowering Plant Family Identification* [20], encontramos un sistema muy similar al anterior con menor flexibilidad, ya que no tiene ningún medio de filtrado ni diferentes tipos de respuesta por característica.

En *An Expert system for integrating several manuals on determining plant diseases into one united database* [22], se describe la normalización de conocimiento existente en los distintos manuales de diagnóstico de enfermedades en plantas. Aunque no es exactamente el mismo tipo de sistema que estamos diseñando. La idea que persigue es

crear un instrumento universal basado en sistemas expertos, capaz de integrar los datos de varios manuales en una sola base de datos unida. Nuestra preocupación en este estudio ha dado el primer paso en este sentido, cómo representar la información. En un futuro se podría ampliar por medio de ejemplos e incluso sistemas inteligentes que pudieran extraer información de las descripciones.

En la documentación revisada hemos podido conocer un estudio *Modelos de representación del conocimiento para la identificación taxonómica y aplicaciones* [23][24], que utiliza técnicas de inteligencia artificial para abordar las problemáticas de un sistema de determinación. Este estudio sigue unas pautas y preocupaciones similares a las nuestras y a las de cualquier investigación que quiera enfrentarse a la problemática de la determinación taxonómica. Comentaremos a continuación las principales aportaciones detectadas. La autora nos explica la creación de un sistema para la identificación taxonómica a través de un sistema experto. Además, presta especial interés a la representación taxonómica del conocimiento y la búsqueda de un repositorio común de información sobre el que trabajar. Finalmente, no solo realiza un sistema para la determinación, también desarrolla una herramienta que permite la generación de claves dicotómicas por medio de ejemplos. En el futuro, como veremos, nuestro sistema podría evolucionar y contener aprendizaje con ejemplos.

Este estudio es muy interesante ya que ha podido recoger resultados de experimentos realizados con botánicos. Adicionalmente, la tesis recoge una amplia documentación a cerca de la determinación taxonómica, diferentes investigaciones realizadas en este ámbito y la descripción de algunos sistemas conocidos de determinación más conocidos que pueden servir de referencia para futuras investigaciones.

2.4. Estudio previo. Prímulas

Previamente a este estudio, se realizó un pequeño proyecto de determinación de la familia de las Prímulas que pretendía adentrarse en el ámbito de la determinación taxonómica mediante lógica borrosa. De las descripciones y claves conocidas de la familia Prímulas, se determinaron los caracteres o atributos que lo especifican: color de la corola, flores en umbela o aisladas, forma del cáliz, forma y tamaño de las hojas, etc. Para cada uno de ellos se estudió su naturaleza, algunos de ellos se mostraron como atributos medibles y otros no. Los tipos detectados fueron: booleano (o de respuesta nítida), numérico medible en un intervalo conocido, borroso o interpretable cuando aparece falta de nitidez o información. Para cada atributo, se establecieron los distintos valores que podían tomar de la propia literatura de la Flora y se atribuyeron etiquetas lingüísticas para cada atributo y especie. A través de conjuntos borrosos, se procedió al

desarrollo de reglas borrosas para cada uno de ellos que permitieran definir de forma correcta cuales son las características morfológicas de cada una de las especies.

El estudio fue un primer experimento a fin de validar la utilidad de la lógica borrosa en el campo de la identificación botánica. Las dificultades de este estudio fueron, la dificultad del vocabulario botánico, que se ha superado en esta ocasión gracias a la participación de un experto botánico. Así como, la generalización de características y sus correspondientes conjuntos borrosos.

En la Figura 2-8, podemos observar las características que fueron modeladas, booleanas, numéricas y borrosas; identificadas respectivamente por un desplegable con las opciones disponibles, una caja de introducción de texto para las numéricas y una escala para las borrosas. En la parte inferior aparecen las 8 posibles especies cada una con el grado de pertenencia inferido para cada una de ellas calculado a partir de dos t-normas, Mínimo y Producto.

Lógica Fuzzy. Sistema de Reconocimiento de Primulas.

Entradas del Sistema

Farinosa	farinosa	<input checked="" type="radio"/>	Inflorescencia	una flor = aisladas ————— muchas flores	<input type="radio"/>
Hoja			Limbo		
Largo Hoja	4.00	<input checked="" type="radio"/>	Máxima Anchura	inf ————— sup	<input checked="" type="radio"/>
Ancho Hoja	00.00	<input type="radio"/>	Borde	Denticulado no dentado = entero ————— muy dentado <input type="radio"/>	<input type="radio"/>
Pecíolo			Crenado	no crenado = entero ————— muy crenado	<input checked="" type="radio"/>
Largo Pecíolo	00.00	<input type="radio"/>	Rugoso	no rugoso ————— muy rugoso	<input checked="" type="radio"/>
Pecíolo Alado	si ————— no	<input type="radio"/>	Base	atenuado ————— bruscamente contraído	<input type="radio"/>

Salidas del Sistema

Nombre	Elatior	Veris	Acaulis	Farinosa	Integrifolia	Latifolia	Pedemontana	Hirsuta
Mínimo	0,50	0,00	0,20	0,00	0,75	0,00	1,00	1,00
Producto	0,38	0,00	0,05	0,00	0,75	0,00	1,00	1,00

Figura 2-8 Pantalla con la interface gráfica del programa de identificación de Primulas.

Capítulo 3

Material y métodos

3.1. Lógica borrosa

En el Apéndice1 pag 189 podemos encontrar una introducción a la lógica borrosa, si bien se detallan en este capítulo algunos conceptos básicos que nos servirán de apoyo para los sistemas borrosos.

3.1.1. Conjuntos borrosos

La teoría de los conjuntos borrosos forma parte del ámbito de la “Inteligencia Artificial”. La cual investiga cómo hacer que los ordenadores puedan realizar las mismas tareas que las mentes humanas. La teoría clásica de conjuntos de *Cantor* no es capaz de recoger aquellos fenómenos reales cuyas características son "imprecisas" o "difusas". En la esfera de los predicados subjetivos, y por tanto imprecisos, la teoría de conjuntos clásica se enfrenta con obstáculos difíciles de superar. La lógica clásica, la teoría de conjuntos clásica o la teoría de probabilidad no siempre resultan adecuadas para tratar la imprecisión, la incertidumbre, la no especificidad, la vaguedad, la inconsistencia y la complejidad del mundo real. Esto motiva la evolución de los conjuntos difusos y las lógicas borrosas y explica su papel en la reestructuración de los fundamentos de las teorías científicas y sus aplicaciones.

Lotfi A. Zadeh en su artículo “*Fuzzy Sets*” [6] introduce una teoría sobre unos objetos, los conjuntos difusos, que son conjuntos de frontera no precisa y cuya función de pertenencia indica un grado. En dicho artículo crea la base teórica sobre subconjuntos borrosos y de la que parten todas las investigaciones posteriores sobre el tema.

La teoría de subconjuntos borrosos surge debido a querer trabajar con la imprecisión, la incertidumbre y la vaguedad, pretendiendo eliminar estas limitaciones habituales que encontramos en la teoría clásica de conjuntos, haciendo de ésta un caso particular de la anterior. La lógica queda modificada, ya no es una lógica bivalente en la que si algo es cierto su contrario es necesariamente falso.

3.1.2. Lógica borrosa

El desarrollo de la tecnología computacional ha abierto diversos campos de investigación. Se pretende que una máquina, o sistema de ellas, pueda producir razonamientos o acciones que, si fuesen realizados por una persona, serían considerados inteligentes. El intento de automatizar el razonamiento y el aprendizaje requiere trabajar sobre las lógicas borrosas.

Las lógicas borrosas necesitan generalizar las conectivas entre conjuntos borrosos. Los conectivos lógicos AND, OR y NOT, y las operaciones entre conjuntos intersección, unión y negación, se generalizan respectivamente mediante normas triangulares, conormas triangulares y negaciones.

En este trabajo se han utilizado tres lógicas fundamentales para los operadores AND y OR, presentes en los antecedentes de las reglas: la t-norma mínimo y t-conorma máximo (de la lógica de Zadeh), t-norma producto y t-conorma suma-producto y t-norma operación y t-conorma operación dual de Lukasiewicz.

3.1.3. Etiquetas lingüísticas

Según L. A. Zadeh, un aspecto clave de la computación con palabras [8], [9] es que supone una fusión del lenguaje natural y de la computación con variables borrosas.

La teoría de los conjuntos borrosos está implícitamente relacionada con el lenguaje natural mediante el concepto de variable lingüística.

Las variables lingüísticas [7] son variables cuyos valores se representan mediante términos lingüísticos. El significado de estos términos lingüísticos se determina mediante conjuntos difusos.

Es especialmente notable la capacidad de las variables lingüísticas para manipular percepciones (de distancias, tamaños, pesos, color...), necesarias en un problema de determinación de plantas.

La diferencia esencial entre percepciones y medidas reside en que las medidas son nítidas (crisp) y las percepciones borrosas (fuzzy). Esta es la razón fundamental por lo que el sistema lógico empleado debe ser fuzzy.

Por ejemplo, dada una característica A , podemos modelar 3 percepciones ' x es poco A ', ' x es medianamente A ' y ' x es muy A ' a partir de una variable sobre x mediante los conjuntos borrosos de la siguiente figura

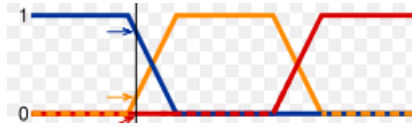


Figura 3-1. Ejemplo de modelo de variables lingüísticas

Dada una variable lingüística ' x es A ', cuyo significado es determinado por un conjunto borroso, es fácil definir otras variables lingüísticas como ' x es poco A ', ' x es muy A ', ' x es más o menos A ', mediante conjuntos borrosos obtenidos a partir de modificadores lingüísticos.

3.1.4. Inferencia y razonamiento aproximado

Cuando se pretende utilizar las relaciones borrosas para efectuar inferencias de razonamiento aproximado de la forma

$$\begin{array}{l} \text{Si } x \text{ es } P \text{ entonces } y \text{ es } Q \\ x \text{ es } P' \\ \hline y \text{ es } Q' \end{array}$$

se puede utilizar la regla composicional de inferencia [2] entre un conjunto borroso que represente la etiqueta lingüística P en todos los elementos del universo en discurso y una relación borrosa que representa la regla *Si ' x es P ' entonces ' y es Q '*, que puede ser generada por diferentes operadores de implicación [3], como Mandami, implicaciones residuadas, S-implicaciones, QM-implicaciones, etc.

3.2. XFUZZY

Se ha utilizado como software la herramienta Xfuzzy por sus diversas funcionalidades.

Esta herramienta fue diseñada e implementada en la Universidad de Sevilla, como un entorno de desarrollo de sistemas de inferencia basados en lógica difusa. Está formada por diversas herramientas que cubren las diferentes etapas del proceso de diseño de un sistema de inferencia difuso.

Entre sus características más notorias se encuentran la capacidad para el desarrollo de sistemas complejos y la flexibilidad para extender el conjunto de funciones difusas disponibles en la herramienta. Xfuzzy ha sido programado en Java y tiene licencia GNU.

Xfuzzy está dividido en diferentes herramientas que pueden ser ejecutadas como programas independientes. El entorno las integra bajo una misma interfaz gráfica que facilita a los usuarios el proceso de diseño. En el flujo de diseño de Xfuzzy estas herramientas se encuentran divididas por etapas: (Figura 1).

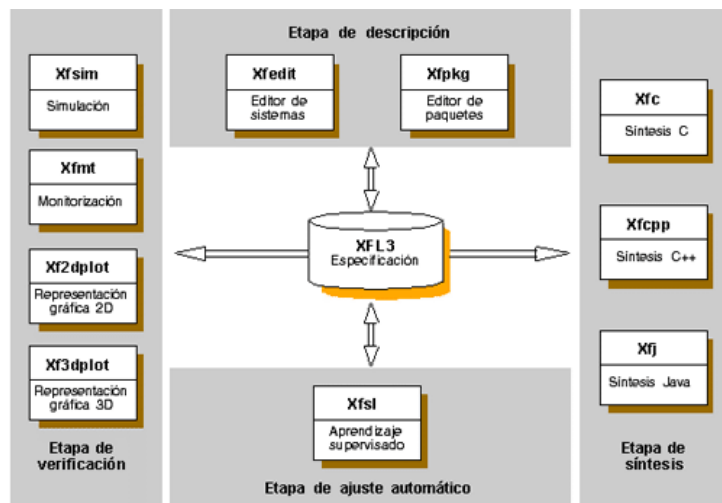


Figura 3-2. Flujo de Diseño de Xfuzzy 3.0.

- La etapa de descripción incluye herramientas gráficas para la definición de sistemas difusos y la definición de paquetes de funciones.
- La etapa de verificación está compuesta por herramientas para la simulación, monitorización y representación gráfica del comportamiento del sistema.
- La etapa de ajuste facilita la aplicación de algoritmos de aprendizaje a los sistemas ya diseñados.
- La etapa de síntesis incluye herramientas para generar descripciones en lenguajes de alto nivel (C, C++ y Java) para implementaciones software o hardware sobre los sistemas de inferencia difusos diseñados.

Xfuzzy utiliza un lenguaje de especificación que es común a todas las herramientas, y que sirve de nexo entre ellas, el XFL (XFuzzy Language). Es un lenguaje flexible y potente, que permite expresar relaciones muy complejas entre variables difusas por medio de bases de reglas jerárquicas y conectivas, modificadores lingüísticos, funciones de pertenencia y métodos de defuzzificación definidos por el usuario.

Capítulo 4

Determinación taxonómica. Conocimiento

La determinación taxonómica de especímenes dentro del campo de la botánica se compone de una serie de observaciones y evaluaciones realizadas por el experto, que se convierte en el usuario del sistema de determinación. Ilustrativamente podemos ver como el usuario en el método clásico Dicotómico se ve guiado, a través de una serie de preguntas, a observar y finalmente evaluar la respuesta a su juicio más adecuada a esa pregunta. Si el usuario utilizase un sistema genérico cualquiera de determinación seguiría este mismo patrón, que consta de una serie de características apreciables en el objeto de estudio sobre las cuales realizar una valoración subjetiva respecto del conocimiento y visión del usuario sobre las circunstancias concretas del objeto. De esta manera, el experto establece predicados estimados que conducen a una identificación del objeto, estos predicados llevados a cabo a través de la observación son producto de la situación concreta del objeto y la percepción del usuario, por lo que podemos decir que la conclusión obtenida se convierte en una implicación aproximada.

Podemos ver gráficamente en la siguiente Figura 4-1, cómo interactúan los diferentes elementos hasta derivar en la determinación.



Figura 4-1. Esquema ilustrativo general de los elementos involucrados en la determinación.

El sistema debe proporcionar al usuario una serie de procesos que faciliten la tarea de obtención de la determinación. Cada uno de estos procesos cuentan a modo genérico con tres elementos básicos: observar una característica concreta del objeto de estudio, evaluar haciendo uso del conocimiento la característica en el espécimen, establecer una conclusión concreta sobre la característica.



Figura 4-2. Proceso genérico de sistemas de determinación.

El sistema de determinación debe resolver varias problemáticas. En primer lugar es necesario definir el conocimiento del que se servirá el sistema para establecer los procesos de reconocimiento. Posteriormente se definen dichos procesos, los cuales guiarán al usuario en la identificación, y se fijarán finalmente los métodos de inferencia que establezcan la conclusión.

El conocimiento se compone de dos grandes componentes, la jerarquización taxonómica de la flora a representar junto con el conjunto de características a evaluar. El conocimiento debe establecer las relaciones entre ambas componentes. Para cada una de las características que se identifican se establece el modo de representación de las mismas, su naturaleza y los diferentes modelos de evaluación. La relación intrínseca con la jerarquía taxonómica es compleja, no es una simple asociación, estas relaciones forman parte de la evaluación y posterior conclusión de toda identificación.

Los procesos involucrados en las acciones que el usuario realiza al interactuar con el sistema, son aquellos que definen qué características tiene que observar el usuario, cómo debe determinar su evaluación y cómo actúa el sistema cuando recibe la conclusión. Por ejemplo, el método Dicotómico clásico, presenta una serie de

características a modo de preguntas, el método de evaluación que proporciona es una respuesta con únicamente dos posibilidades (si o no) y un único camino tras recibir la conclusión.

La información que nuestro sistema utiliza como conocimiento no tiene como base repositorios de información conocidos, si no que se ha llevado a cabo una investigación y definición de un modelo propio desarrollado con la información recogida del apartado *Gymnospermas* del *Tomo I* de Flora Ibérica [10]. La información se ha extraído paralelamente a partir de información contenida en las claves de Flora Ibérica y en las descripciones de las fichas de cada uno de los taxones.

La clasificación que se ha realizado incluye un subconjunto de Flora Ibérica, el sistema es adaptativo, como se comentará más adelante en apartado 4.4.1. Está preparado para incluir cualquier modificación y/o añadidura automáticamente sin necesidad de ningún tipo de cambio de código en el programa.

La evaluación técnica botánica se realiza sobre la documentación mencionada, obteniendo diversos elementos del modelo que nos permitirán diseñar el sistema de inferencia. Se irán comentando en este apartado.

4.1. Conocimiento. Flora (Taxones).

En primer lugar el conocimiento se compone de la información jerarquizada de los taxones pertenecientes a la Flora Ibérica. Esta información consiste en una estructuración de los elementos atendiendo al tipo de taxón de cada uno y sus taxones “hijo”. De tal manera que las familias tendrán una serie de hijos del tipo de taxón Género; los géneros tendrán hijos de tipo de taxón especie y así sucesivamente. Tal y como se ilustra en la Figura 4-3.

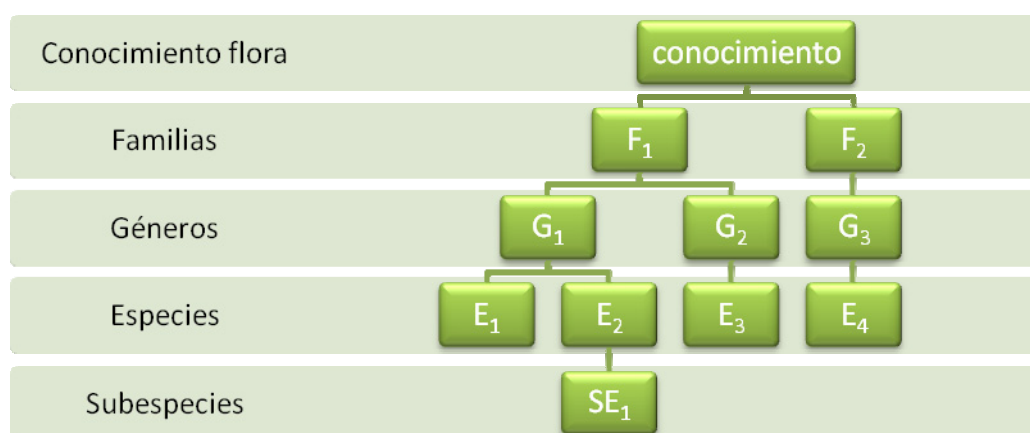


Figura 4-3. Esquema jerárquico de la estructuración del conocimiento de Flora Ibérica.

4.2. Conocimiento. Características.

Del estudio de las características encontradas en la definición de los distintos taxones, se extrae la cualificación de dichas características y sus asociaciones para obtener los elementos de evaluación del sistema.

4.2.1. Agrupaciones de características

Tras un minucioso estudio se seleccionan las características representativas para la identificación y se establece una clasificación las mismas generando **agrupaciones** de características, atendiendo esta clasificación a sus cualidades comunes. Estas agrupaciones reciben un nombre determinado y pasan a ser un elemento más utilizado en la estructuración del conocimiento. Algunos ejemplos de agrupaciones son:

- *Planta*
- *Biotipo*
- *Forma de la Copa*
- *Tallo*
- *Tronco*
- *Ritidoma (parte exterior de la corteza, que se desprende)*
- *Ramillas (brotes del año)*

Una vez recopiladas las agrupaciones se detecta si existen asociaciones entre agrupaciones y características, distinguiendo aquellas características que tienen que ser evaluadas positivamente para que dicha agrupación tenga sentido. En el ejemplo siguiente vemos dos agrupaciones *Biotipo* y *Forma de la Copa*, no tiene sentido hablar de la forma de la copa si el espécimen de estudio no es un árbol, por lo tanto la agrupación *Forma de la Copa* está intrínsecamente relacionada con la característica *Árbol (fanerófito)*.

- Agrupación: *Biotipo*
 - Características pertenecientes a la agrupación:
 - Árbol (fanerófito) → *1
 - Arbusto (nanofanerófito)
 - Matorral bajo (caméfito)

- Altura (m)
- Agrupación; *Forma de la* Copa → Depende de la característica Árbol (fanerófito)
*1
 - Características pertenecientes a la agrupación:
 - Copa cónica o piramidal
 - Copa redondeada
 - Copa aparasolada, densa.
 - Copa irregular

A este tipo de agrupaciones les llamamos **agrupaciones ocultas**, si pensamos en estructurar la información de las características en las diferentes agrupaciones detectadas, y mostrar esta información en el sistema al usuario; no se debe mostrar una agrupación oculta si su característica asociada no ha sido seleccionada o contestada por el usuario. Adicionalmente una misma característica puede estar asociada a más de una agrupación oculta, por lo tanto la relación de agrupación oculta $AG(C^{[1..1]}, A^{[1..*]})$, se define mediante la característica con multiplicidad [1..1] y sus agrupaciones multiplicidad [1..*].

Este mecanismo nos permite simplificar y estructurar la información solicitada al usuario en el proceso de determinación.

4.2.2. Características Cuantitativas

Si observamos un objeto de estudio para su identificación es fácil detectar características medibles numéricamente, como por ejemplo el número de flores o la altura del espécimen. Estas características que el observador es capaz de medir de forma nítida y objetiva con una representación numérica, son denominadas Características Cuantitativas. El comportamiento común de estas características es disponer de uno o varios límites que restringen los posibles valores para esa característica concreta en un taxón concreto. Existen seis límites posibles detectados en Flora Ibérica para las características, estos son:

Tabla 4-1. Descripción de límites de características cuantitativas.

Max	Máximo
Min	Mínimo
MaxInt	Máximo del intervalo frecuente
MinInt	Mínimo del intervalo frecuente
Med	Media
ES	Error estándar

Para cada una de las características es necesario recopilar información sobre cuáles de estos límites son utilizados para describirla en cada una de las relaciones establecidas con los taxones. Hay que tener en cuenta que no tienen porqué coincidir los límites usados para una misma característica en relación con un taxón llamémosle T_x y otro diferente T_y .

Para cada taxón del que se disponga información relativa a una característica cuantitativa C_x se recopilarán los límites conocidos en la literatura para esta relación, si el taxón seleccionado es T_x hablamos de la relación $R_{xx}(T_x, C_x)$. De esta manera la relación $R1$ se autodefine con los valores adoptados para dichos límites y contiene una lista de pares (límite, valor) definida como $[(L_n, V_n)]$ donde $n \leq 6$.

Si realizamos este mismo proceso para cada uno de los taxones T_i relacionados con la característica C_j , podemos extraer el conjunto que incluye todos los límites establecidos en las relaciones como la lista de límites $[L_n]$ donde $n \leq 6$ y todo límite perteneciente a la lista (límite, valor) de cada una de las relaciones R_{ij} de la característica C_j pertenece a la lista $[L_n]$.

En cada una de las relaciones las características se determinan utilizando un subconjunto de uno de los siguientes dos grandes conjuntos (Min, MinInt, MaxInt, Max) o bien (Med, ES). Si tomamos valores mayores que 0, ambos conjuntos pueden verse representados en una gráfica en el orden que muestra la Figura 4-4 y la Figura 4-5.



Figura 4-4. Gráfica del conjunto de límites (Min, MinInt, MaxInt, Max) para una relación $R_{xx}(T_x, C_x)$.



Figura 4-5. Gráfica del conjunto de límites (Med, ES) para una relación $R_{xx}(T_x, C_x)$.

4.2.3. Características Cualitativas

Las características cualitativas son atributos observables no medibles numéricamente. Típicamente requieren una respuesta positiva o negativa del usuario, por tanto la enunciación de dichos predicados debe adaptarse a conclusiones booleanas. Podemos observar el siguiente ejemplo perteneciente a la agrupación *Tronco*:

- Agrupación: Tronco
 - Tronco profundamente hendido
 - Tronco sólo ramificado en la parte superior
 - Tronco de color pardo rojizo
 - Tronco tortuoso
 - Tronco Recto

En los sistemas clásicos la información que se extrae de la evaluación es limitada, ya que únicamente permite dos posibles conclusiones “sí” y “no”, es decir “lo es” o “no lo es”. Las características observables en cualquier espécimen pueden ser de muy diversas naturalezas, algunas de ellas pueden resultar triviales o genéricas a la hora de proceder a una identificación taxonómica, mientras que otras son características “claves” cuya observación puede inferir la determinación taxonómica con una certeza superior.

Nuestro modelo refleja un nuevo atributo o cualidad asociada a toda relación conocida R_{xx} entre un taxón y una característica $R_{xx}(T_x, C_x)$. Este atributo representa la importancia, en base al grado de posibilidad, de la característica para llevar a cabo una correcta identificación. Nos permite jerarquizar las características relacionadas con un taxón concreto por su nivel de posibilidad y utilizar esta ponderación en la inferencia de la conclusión. Ilustrativamente, si un usuario ha respondido positivamente a cinco características relacionadas con el taxón T_x y otras cinco características relacionadas, iguales o diferentes, con un taxón diferente T_y ; gracias al atributo de importancia asignado a cada una de esas relaciones, podemos determinar el taxón con mayor grado de posibilidad, aunque aparentemente ambos taxones son igualmente posibles por tener cinco características satisfactorias. De tal manera que si el conjunto de relaciones de características afirmativas de T_x contiene aquellas marcadas con importancia superior, mientras que el conjunto de T_y contiene únicamente características de poca importancia. Podemos concluir que las relaciones afirmativas de T_x son más fuertes y ordenar los posibles resultados, priorizando los de mayor posibilidad.

Este atributo de importancia o posibilidad se modela asignando un identificador numérico a los diferentes grados obtenidos en el estudio, de tal manera que el 1 se corresponde con el grado inferior y 3 con el grado superior. Adicionalmente a esta clasificación numérica, se le asocia un predicado concreto para facilitar al experto botánico la tarea de asignar dichos valores a las relaciones (taxón, característica) a la hora de alimentar el conocimiento del sistema. En la Tabla 4-2 podemos encontrar dichos significados.

Tabla 4-2. Posibilidades positivas asociadas a una relación $R_{xx}(T_x, C_x)$.

1	En ocasiones se da. Es posible, aunque no muy frecuente
2	Frecuentemente, normalmente.
3	Siempre se da.

Por lo tanto, toda relación que se detecte entre un taxón y una característica será evaluada por medio del grado de posibilidad de la misma. Esta clasificación formará parte del conocimiento del sistema.

No solo la relación positiva nos proporciona información útil para la determinación, sino también la ausencia de información (no se conoce relación entre taxón y característica) o la información negativa (es imposible que exista relación positiva entre ambas).

Por lo tanto, se marcarán así mismo aquellas parejas (taxón, característica) para las que no dispongamos de información sobre relación alguna entre ambas. Su grado de posibilidad será 0 - No hay información.

Por último, como hemos destacado, existen relaciones negativas cuando una característica C_x es imposible que sea observada en un espécimen cualquiera del taxón T_x . Marcaremos en este caso la posibilidad de $R_{xx}(T_x, C_x)$ igual a -1. De tal manera que si el usuario está observando un espécimen y evalúa la característica C_x como positiva, automáticamente rechazaremos el taxón T_x como parte del conjunto de conclusiones inferidas posibles.

Como sumario derivamos que toda relación $R_{xx}(T_x, C_x)$ tendrá un atributo asociado de posibilidad que se rige por los valores descritos en la Tabla 4-3 junto con sus predicados descriptivos.

Tabla 4-3. Conjunto completo de posibilidades asociadas a una relación $R_{xx}(T_x, C_x)$.

-1	No es posible. Ese taxón quedaría rechazado si se da la característica
0	No hay información
1	En ocasiones se da. Es posible, aunque no muy frecuente
2	Frecuentemente, normalmente.
3	Siempre se da.

El modo en el que el experto botánico identifica estos valores al estudiar las claves y descripciones de los taxones de la Flora Ibérica es el siguiente:

- -1 si el hecho de que se de esa característica elimina a la especie en cuestión.
- 0 si no hay información sobre esa característica para esa especie
- 1 si la especie en cuestión puede tener esa característica, pero es poco frecuente que la tenga. En el lenguaje de la publicación corresponde a "En ocasiones", "algunas veces" o expresiones equivalentes.
- 2 si la especie en cuestión normalmente tiene esa característica, aunque en ocasiones puede no tenerla. Equivale a expresiones como "normalmente", "usualmente", "En general".
- 3 cuando la especie siempre tiene (o cumple) esa característica. Se procede a buscar la característica contraria auto-excluyente que deberá tener un -1.

La información extraída de las claves suele ser la más excluyente, ya que produce gran cantidad de -1 y 3, y pocos 1 y 2. Las descripciones de los taxones son normalmente más difusas y son las que más indeterminación producen con valores 1 y 2.

Por ejemplo, si en la descripción de una especie encontramos: "Árbol.". Significa inequívocamente que siempre es un árbol, por lo que la característica "Árbol" tendrá un 3 y las características "Arbusto" y "Herbácea" tendrán un -1. Estas tres características son auto-excluyentes, no se pueden dar simultáneamente.

Sin embargo si la descripción dice:

"De porte arbóreo, aunque en ocasiones puede ser arbustivo"

Esto significa que "Árbol" tendrá un 2, "Arbusto" un 1 y "Herbácea" un -1.

4.2.3.1. Características cualitativas booleanas

Existen características no medibles que tienen como respuesta únicamente *sí* o *no*, la información que se obtiene es completamente nítida, pues toma grados de pertenencia 0 o 1. Esta es la aproximación que fue diseñada en el estudio preliminar sobre la familia Prímula [11], si bien en este estudio posterior veremos como la nueva representación del conocimiento nos permite extraer información de estas características.

Un ejemplo de estas características es *Árbol (fanerófito)*, solo tiene dos posibles respuestas, un espécimen o es Árbol o no lo es.

4.2.3.2. Características cualitativas aproximadas

En contraposición con las características cualitativas booleanas se detecta la existencia de características con respuestas aproximadas, se trata de cualidades que el usuario puede observar en mayor o menor grado, o incluso puede servirse de esta aproximación si no está seguro de estar observando plenamente dicha característica. Las característica *fusiformes* de la agrupación *Yemas* es un ejemplo de este tipo, el usuario puede utilizar el lenguaje natural para describir la cualidad observada como *poco*, *medianamente* o *muy fusiforme*.

Estas características cobran especial interés como elementos pertenecientes a sistemas borrosos de inferencia, sin embargo en sistemas nítidos se comportan como características booleanas. Es decir, en un sistema nítido si una característica de este tipo es seleccionada en un grado mayor que 0, donde 0 es el límite mínimo, se toma la respuesta como un *sí* booleano. Si la respuesta se selecciona en el mínimo se toma como un *no*.

4.2.3.3. Características auto-excluyentes

Tras disponer de los diferentes conjuntos de características que componen las agrupaciones detectadas en el estudio, podemos analizar posibles relaciones entre estas características. La principal asociación que encontramos es la categorización de **auto-excluyentes**. Existen subconjuntos dentro de las agrupaciones que se componen de características auto-excluyentes, es decir solo podrá ser observada en un espécimen simultáneamente una única característica del conjunto.

Una respuesta positiva a alguna de las características de un conjunto de auto-excluyentes, conllevará automáticamente a la respuesta negativa del resto de elementos del conjunto.

El experto botánico, en la tarea de conformación del conocimiento del sistema, tendrá que identificar y marcar los conjuntos auto-excluyentes; de esta manera se convierte en un elemento más del sistema, involucrado en la inferencia de la identificación taxonómica.

Podemos ver un ejemplo de estas agrupaciones en el siguiente ejemplo, donde un espécimen podrá ser Árbol, Arbusto o Matorral, pero nunca más de dos de estas características a la vez.

- Agrupación: *Biotipo*
 - Subconjunto de características auto-excluyentes:

- Árbol (fanerófito)
- Arbusto (nanofanerófito)
- Matorral bajo (caméfito)

Las características cualitativas pertenecientes a un grupo de exclusión pueden ser booleanas y aproximadas. A continuación vemos un nuevo elemento de determinación investigado.

4.2.3.3.1. Características cualitativas auto-excluyentes relacionadas aproximadas

En los conjuntos de auto-excluyentes podemos encontrar características relacionadas las cuales tienen una “frontera” en la que ambas cualidades podrían darse y el observador podría no estar seguro de elegir una u otra. Como por ejemplo la agrupación *Color de hojas* tiene dos características dentro del grupo de autoexclusión 22, concretamente *Color verde oscuro* y *Color verde claro*. La diferenciación en ocasiones de ambas características puede hacerse dificultosa ya que depende de la percepción del usuario.

Si pensamos en poder dar al usuario un modo de aproximarse a estas características aunque no tenga certeza de cuál de ellas escoger como respuesta. Podemos describir las características de la siguiente manera:

- *Color verde oscuro* → Extremadamente **muy** verde oscuro
- *Color verde claro* → Extremadamente **poco** verde oscuro

Entre el rango que construyen ambas descripciones el usuario puede utilizar el lenguaje natural para acercarse más al *poco* o al *muy*. De esta manera dejamos al usuario aproximarse a la característica que aprecia en mayor medida, sin necesidad de obligarle a elegir nítidamente una de ellas.

Al igual que en las características aproximadas no relacionadas que vimos anteriormente en el punto 4.2.3.2, este nuevo elemento descrito es útil en un sistema borroso de determinación taxonómica, pero carece de importancia en sistemas nítidos; los cuales tomaran el acercamiento a una de las características como la respuesta booleana positiva *sí* (1) y para las características alejadas inferirá respuesta negativa *no* (0).

4.3. Representación numérica del lenguaje científico

Si observamos los elementos estudiados en los dos puntos inmediatamente anteriores, tras el análisis de la información, hemos asignado identificadores y valores numéricos a los elementos detectados como útiles para un sistema de determinación taxonómica.

El conjunto de elementos que hemos obtenido y la posterior representación desarrollada para cada elemento; nos ha permitido, partiendo de un lenguaje científico (botánico) expresado en lenguaje natural, obtener una representación numérica del mismo.

4.4. Representación del conocimiento

4.4.1. Sistema adaptativo

Para asegurar la utilidad en el tiempo del sistema, se establece como requisito diseñar la alimentación de conocimiento de tal manera, que sea posible la modificación de datos ya introducidos y la inserción de nuevo conocimiento sin necesidad de modificar el código programático.

El sistema necesita una primera alimentación de conocimiento básico, que actualmente se ha trabajado con un experto botánico. Las ventajas que ofrece la adaptabilidad es que este conocimiento no solo podría ser incrementado por un único experto en propia distribución de software, si no que podría intercambiar su información con otros expertos y convertirse en una retroalimentación continua de conocimiento, pudiendo alcanzar una base de información sólida y de gran completitud. En el futuro se podría diseñar un procedimiento de inserción y comprobación de los datos introducidos por los distintos expertos en una aplicación web, y ofrecer descargas de actualizaciones periódicas de dicho conocimiento, el cual se actualizaría automáticamente en la distribución personal de cada usuario.

La información contenida en estas tablas, será traducida a documentos XML como se verá más adelante. Inicialmente se estudió realizar una representación jerárquica tipo XML para la introducción de datos por parte del experto. Pero tras la toma requisitos con el experto botánico, se desestimó ya que los usuarios habituales pueden no estar familiarizados con este tipo de representaciones.

El sistema debe de adaptar tanto su visualización como su lógica interna a cada nueva incorporación que se realice. Se actualizará la interface gráfica para ser capaz de mostrar la información al usuario, todas las relaciones implicadas en las nuevas

incorporaciones. La evaluación de las premisas tendrá también que incluir esta información así como las reglas de inferencia que obtengan el conjunto solución final.

Esta adaptabilidad sin límites, podría evolucionar de tal manera que permitiera generalizar las características para que el sistema pudiera funcionar como un 'determinador universal'. Es decir, el sistema permitiría como entrada cualquier problemática de determinación (plantas, coches, animales...) y, tras la generalización de las reglas de inferencia, conseguiría establecer un conjunto de solución.

4.4.2. Introducción de datos

El objetivo de un sistema adaptativo es permitir al usuario introducir modificaciones y novedades al programa de forma autónoma. Es necesario ofrecer al usuario la representación de la información y los mecanismos necesarios para aumentar el conocimiento del mismo.

Tras realizar un análisis de requisitos con el experto botánico, determinamos el uso de una hoja de cálculo como herramienta inicial de introducción de datos, medio conocido para los usuarios finales. Una futura evolución del sistema podría incluir un asistente gráfico estructurado para proceder a la correcta inserción de la información de una forma más controlada. El método adoptado actualmente no ha supuesto complicaciones reseñables, si bien sería recomendable evolucionar al asistente comentado.

La estructura de la tabla es compleja, esta complejidad bien motivada por tener que contener todos los elementos que intervienen en el conocimiento que hemos comentado.

4.4.2.1. Flora (Taxones)

En primer lugar disponemos de una lista de taxones, pertenecientes a la Flora Ibérica, que debemos estructurar. Como hemos visto esta estructura se realiza utilizando el tipo de taxón y jerarquizando posteriormente los taxones hijos de cada elemento.

Utilizaremos el siguiente ejemplo reflejado en la

para explicar la representación elegida.

Tabla 4-4 Ejemplo de estructuración de taxones.

	PINACEAE	Abies	Abies pinsapo	Abies alba	Pinus	Pinus halepensis	Pinus uncinata	Pinus pinea	Pinus radiata	Pinus sylvestris	Pinus nigra	Pinus pinaster	Pinus nigra ssp. nigra	Pinus nigra ssp. salzmannii	CUPRESACEAE	Tetraclinis	Tetraclinis articulata
TIPO DE TAXÓN	1	2	3	3	2	3	3	3	3	3	3	3	4	4	1	2	3
Número de familia	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2
Número de género	0	1	1	1	4	4	4	4	4	4	4	4	4	4	0	8	8
Número de especie	0	0	2	1	0	11	15	10	9	14	13	12	13	13	0	0	22

Los taxones se disponen en columnas, como vemos en el ejemplo, introduciendo ordenadamente familia (verde), género (amarillo), especie (sin color), subespecie (azul).

Los tipos de taxón contemplados en el sistema son los reseñados en la Figura 4-6. A cada uno de los tipos estudiados se le ha asignado un identificador numérico, este identificador es el que podemos ver en la primera fila. De esta manera observamos en la Tabla 4-4 dos familias PINACEAE y CUPRESACEAE con valor 1, y tres géneros *Abies*, *Pinus* y *Tetraclinis* con valor 2.

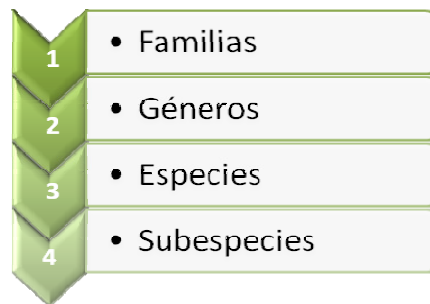


Figura 4-6. Tipos de taxón.

Las siguientes 3 filas, número de familia género y especie, representan la relación jerárquica de los taxones mediante la utilización de las reglas recogidas en la

Tabla 4-5. Estas reglas no son más que asignar un identificador a cada taxón (excepto subespecies), y utilizar dicho identificador para relacionar los elementos hijos en cada caso particular. La Figura 4-7 muestra a modo de gráfico jerárquico las asociaciones establecidas por los identificadores al representar la estructura.

Tabla 4-5. Reglas a utilizar para establecer la relación jerárquica de taxones.

	Familia	Genero	Especie	Subespecie
Número de familia	Identificador de la familia F_i	Identificador de la familia F_j a la que pertenece el género G_j	Identificador de la familia F_k a la que pertenece el género G_k	Identificador de la familia F_m a la que pertenece el género G_m
Número de género	N/A, valor 0	Identificador propio del género G_j	Identificador de la género G_k al que pertenece la especie E_k	Identificador del género G_m al que pertenece la especie E_m
Número de especie	N/A, valor 0	N/A, valor 0	Identificador de la especie E_k	Identificador de la especie E_m a la que pertenece la subespecie SE_m

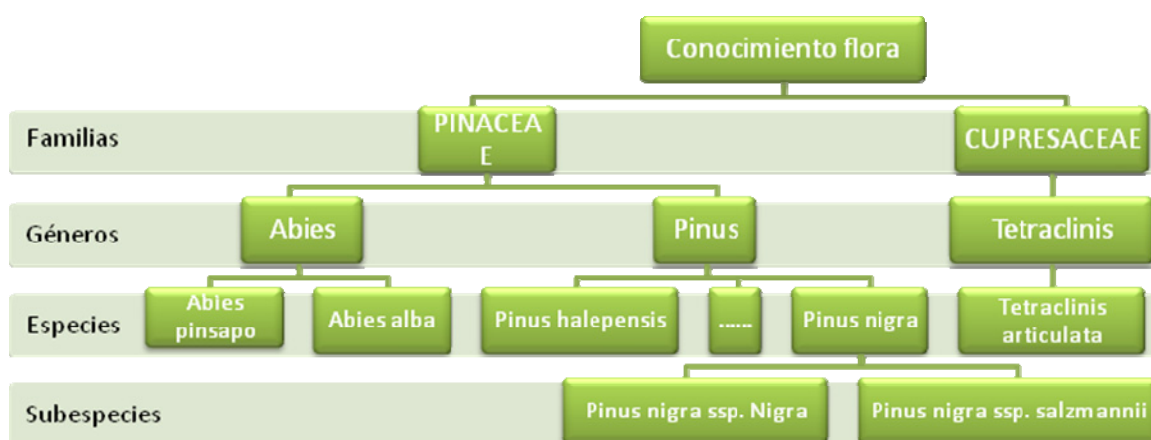


Figura 4-7. Jerarquía representada en el ejemplo de la Características

La representación de las características describe cada uno de los elementos del conocimiento que se detectaron en la primera fase del estudio.

Agrupaciones de características

Las características utilizarán dos columnas, una primera columna para la agrupación y una segunda columna para la lista de características que componen la agrupación.

Tabla 4-6. Ejemplo de estructuración de agrupaciones.

Agrupación	Características
Planta	
	Leñosa
	Herbácea
Forma de la Copa	
	Copa cónica o piramidal
	Copa redondeada
	Copa aparasolada, densa.
	Copa irregular

Agrupaciones ocultas

Las agrupaciones ocultas son aquellas que dependen de que el usuario evalúe positivamente una característica a la que está asociada la agrupación, como ya comentamos en el punto 4.2.1. La representación de esta cualidad se realiza por medio de un identificador que representa la asociación entre la característica y la agrupación o agrupaciones ocultas. Recordemos que una misma característica puede estar asociada a más de una agrupación. Este identificador se recoge en una nueva columna *Agrupaciones ocultas* en la cual se registra el identificador de la asociación junto al nombre de la característica y repetido junto al nombre de cada una de las agrupaciones que están asociadas a la misma.

Observando el ejemplo de la Tabla 4-7 vemos como la característica *Leñosa* tiene el identificador 2 de asociación con preguntas ocultas y a su vez la agrupación *Biotipo* tiene dicho identificador. Es decir, solo cuando el usuario evalúe positivamente la característica leñosa se le mostrará al usuario la agrupación Biotipo. Del mismo modo la característica *Árbol (fanerófito)* tiene el identificador 1 que asocia dos agrupaciones ocultas *Forma de la Copa* y *Tronco*

Tabla 4-7. Ejemplo de agrupaciones ocultas.

Orden	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas
6		Características de la planta a determinar, cualitativas y cuantitativas		
7	Planta			
8		Leñosa		2
9		Herbácea		
10	Biotipo			2
11		Árbol (fanerófito)		1
12		Arbusto (nanofanerófito)		
13		Matorral bajo (caméfita)		
14		Altura (m)		
15			Max	
16	Forma de la Copa			1
17		Copa cónica o piramidal		
18		Copa redondeada		
19		Copa aparasolada, densa.		
20		Copa irregular		
26	Tronco			1
27		Tronco profundamente hendido		
28		Tronco sólo ramificado en la parte superior		
29		Tronco de color pardo rojizo		
30		Tronco tortuoso		
31		Tronco Recto		

Relaciones entre características y taxones

Las relaciones que describimos anteriormente $R_{xx}(T_x, C_x)$, se ven representadas en una tabla, las columnas corresponden a los diferentes taxones y las filas a las características.

En las siguientes dos tablas Tabla 4-8 y Tabla 4-9, podemos ver esquemáticamente la estructura para características cualitativas y cuantitativas respectivamente.

Tabla 4-8. Representación gráfica de las relaciones para características cualitativas.

	T_1	T_2	T_3	T_4
C_1	R_{11}	R_{12}	R_{13}	R_{14}
C_2	R_{21}	R_{22}	R_{23}	R_{24}
C_3	R_{31}	R_{32}	R_{33}	R_{34}
C_4	R_{41}	R_{42}	R_{43}	R_{44}

Tabla 4-9. Representación gráfica de las relaciones para características cuantitativas.

		T_1	T_2	T_3	T_4
C_1					
	L_1	R_{11}	R_{12}	R_{13}	R_{14}
	L_2	R_{21}	R_{22}	R_{23}	R_{24}
C_2					
	L_3	R_{31}	R_{32}	R_{33}	R_{34}
	L_4	R_{41}	R_{42}	R_{43}	R_{44}
	L_5	R_{51}	R_{52}	R_{53}	R_{54}

Los campos con las relaciones R_{ij} , podrán estar vacíos considerándose ausencia de información. Si contienen valores estos serán numéricos. Las características cualitativas se evaluarán según vimos en la Tabla 4-3 con un valor entre (-1, 0, 1, 2 y 3), mientras que las cuantitativas tendrán un valor numérico mayor que 0, asociado a los límites establecidos. La Tabla 4-9 vista anteriormente describe los límites L_x asociados a características cuantitativas.

A continuación veremos más detalladamente cada uno de los tipos de características junto con ejemplos concretos de cada casuística.

Características cuantitativas

Como vimos en el apartado anterior, las características cuantitativas tienen como atributos los límites utilizados para describir los intervalos de conformidad de dichas

características. Fijamos anteriormente 6 posibles valores (Max, MaxInt, Min, MinInt, Med, ES). Los límites posibles asociados a una característica cuantitativa se recogerán en una nueva columna junto a la misma (*Tipo dato Cuantitativo*).

Adicionalmente para mayor facilidad, se introduce una nueva columna que marca las características cuantitativas ("*c*" *cuantitativa*, " " *cualitativa*), con el fin de distinguirlas y realizar su procesamiento particular.

Observando el ejemplo de la Tabla 4-10, tenemos tres características cuantitativas (marcadas en la primera columna con una "*c*"), dos de ellas *Anchura de las hojas (mm)* y *Número de canales resiníferos* tienen como posibles límites Max y Min, mientras que *Longitud de las hojas (mm)* dispone de tres límites.

Tabla 4-10. Ejemplo de características.

Orden	"c" cuantitativa, " " cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	PINACEAE	Pinus	Pinus halepensis	Pinus uncinata	Pinus pinea	Pinus radiata	Pinus sylvestris	Pinus nigra	Pinus pinaster	Pinus nigra ssp. nigra	Pinus nigra ssp. salzmannii
6			Características de la planta a determinar, cualitativas y cuantitativas													
95			Forma de las hojas		4											
96			Hojas escuamiformes			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
97			Hojas aciculares			3	3	3	3	3	3	3	3	3	3	3
98			hojas lineares aplanadas			1	1	1	1	1	1	1	1	1	1	1
99			hojas de sección tetragonal o subtetragonales			1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
100			agudas y/o mucronadas			1	1	0	3	0	0	3	0	3	0	0
101			ligeramente emarginadas			1	0	0	0	0	0	0	0	0	0	0
102			no emarginadas			1	0	0	0	0	0	0	0	0	0	0
103	c		Longitud de las hojas (mm)													
104				Max			250	150	80	200	##	70	160	250	##	##
105				Min			30	60	40	100	70	30	60	100	80	60
106				MaxInt				120	150							
107	c		Anchura de las hojas (mm)													
108				Max				1	2	1,5	2	2	2	2	2	2
109				Min				0,7	1	1	1	1	0,8	1	2	1
122			Canales resiníferos		5											
123			canales resiníferos marginales o submarginales			1	0	3	0	0	0	3	-1	-1	-1	-1
124			canales resiníferos centrales			1	1	-1	0	0	0	-1	3	3	3	3
125	c		Número de canales resiníferos													
126				Max				8					17	18	17	17
127				Min				3					3	2	3	3

Características cualitativas

Observando la estructura genérica descrita en la Tabla 4-8 y el ejemplo concreto de la Tabla 4-10, encontramos características cualitativas y cada una de sus relaciones con los taxones seleccionados. El ejemplo presenta una única familia PINACEAE, y una de sus familias *Pinus* con valores (-1, 0, 1, 2 y 3) asignados a las relaciones.

Características cualitativas booleanas

Las características booleanas por ser las más básicas, en lo que a complejidad de respuestas se refiere, serán identificadas por la ausencia de marca. Son todas aquellas características cualitativas que no dispongan de marca en la columna *Aproximadas*. En los sistemas nítidos que estudiaremos, todas las características cualitativas se comportan como booleanas. La diferenciación es por tanto útil para sistemas de determinación que utilizan lógica borrosa ya que, como veremos, permiten al usuario evaluar las características con cierta aproximación.

Características cualitativas aproximadas

Las características que se detecten como aproximadas serán marcadas en una nueva columna llamada *Aproximadas* con una s. En la Tabla 4-11 podemos observar la representación de estos dos últimos elementos.

Tabla 4-11. Ejemplo de características cualitativas booleanas y aproximadas.

Orden	"C" cuantitativa, "c" cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Aproximadas
80		Yemas				
83			fusiformes			s
86			con escamas recurvadas en la punta			
87			con escamas aplicadas			

Características auto-excluyentes

Estas características forman conjuntos de auto-exclusión en los cuales solo una de las características puede ser observada simultáneamente para un espécimen concreto. Para representar estos conjuntos se ha creado un identificador de conjunto de auto-exclusión que enlaza todas aquellas características que pertenecen al grupo. En la tabla se representa en una nueva columna en la cual se coloca el identificador del conjunto con cada una de las características incluidas en él. Podemos observar el ejemplo de la Tabla 4-12, el grupo de auto-exclusión número 7 contiene cuatro características.

Tabla 4-12. Ejemplo características auto-excluyentes.

Orden	"c" cuantitativa, " " cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Grupo auto-excluyentes	PINACEAE	Abies	Abies pinsapo	Abies alba	Pseudotsuga	Pseudotsuga menziesii	Picea	Picea abies
6			Características de la planta a determinar, cualitativas y cuantitativas											
7			Planta											
8			Leñosa		2	15	3	3	3	3	3	3	3	3
9			Herbácea			15	-1	-1	-1	-1	-1	-1	-1	-1
10			Biotipo		2									
11			Árbol (fanerófito)		1	1	3	3	3	3	3	3	3	3
12			Arbusto (nanofanerófito)			1	-1	-1	-1	-1	-1	-1	-1	-1
13			Matorral bajo (caméfito)			1	-1	-1	-1	-1	-1	-1	-1	-1
14 c			Altura (m)											
15				Max				50	30	50	100	100	60	60
39			Ramificación		1									
40			Ramificación monopódica			7	3	3	3	3	3	3	3	3
41			Ramificación simpódica			7	-1	-1	-1	-1	-1	-1	-1	-1
42			Ramificación verticilada regular			7	1	3	3	3	-1	-1	-1	-1
43			Ramificación verticilada irregular			7	1	-1	-1	-1	3	3	-1	-1
44			Guía y puntas de las ramas recurvadas, colgantes			8	1	0	0	0	0	0	0	0
45			Guía erecta y puntas de las ramas no recurvadas			8	1	0	0	0	0	0	0	0

Características auto-excluyentes relacionadas aproximadas

Para representar estas características, asociaremos un identificador a la relación aproximada existente entre ellas, este identificador le situamos en la columna *Aproximadas* ya existente.

Las relaciones detectadas pueden tener dos o más elementos en el conjunto de características. Si recuperamos el ejemplo que vimos de la agrupación *Yema*, si tuviéramos una nueva característica *Color verde entre claro y oscuro*, tendríamos un grado intermedio. La posibilidad de encontrar este tipo de relaciones existe aunque dado el limitado conjunto de características estudiado, no se ha encontrado ningún ejemplo real concreto.

- *Color verde oscuro* → Extremadamente **muy** verde oscuro
- *Color verde entre claro y oscuro* → **Medianamente** verde oscuro
- *Color verde claro* → Extremadamente **poco** verde oscuro

Para representar la graduación entre las características de la relación es necesario definir una ordenación de estas características, en la Tabla 4-13 podemos observar como el grupo de auto-excluyentes número 18, tiene dos características que pertenecen

a un conjunto de relación de aproximadas número 3. El orden descrito para estas características conlleva a la siguiente descripción:

- Orden 2: *obtusas* → Extremadamente **muy** obtusas
- Orden 1: *agudas* → Extremadamente **poco** obtusas

Mientras que el descrito para la relación de aproximadas 4 es:

- Orden 2: *resinosas* → Extremadamente **muy** resinosas
- Orden 1: *no resinosas* → Extremadamente **poco** resinosas

Tabla 4-13. Ejemplo de características auto-excluyentes relacionadas aproximadas

Orden	"C" cuantitativa, " " cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Grupo auto-excluyentes	Aproximadas	Orden relación proximadas
80		Yemas						
81			agudas			18	3	1
82			obtusas			18	3	2
83			fusiformes			s	s	
84			resinosas			19	4	2
85			no resinosas			19	4	1
86			con escamas recurvadas en la punta			20		
87			con escamas aplicadas			20		

4.4.3. Traducción a metalenguaje estructurado, XML

La herramienta que utiliza el botánico experto, para introducir nuevo conocimiento al sistema, es una tabla Excel a fin de facilitar la tarea al usuario y la adaptación a nuevos sistemas informáticos. Aún así, el sistema no utiliza directamente el Excel para la operativa que realiza. En una primera estancia y siempre que el usuario lo desea, el sistema carga la información de la tabla para transformarla a XML, gracias a los esquemas correspondientes que describen la estructura del conocimiento. De esta manera la información queda almacenada en un repositorio de XMLs. El objetivo es tener una representación final de la información bajo este estándar de intercambio estructurado de datos, que nos permitirá mayor flexibilidad a la hora de evolucionar el sistema, comunicarle con otros sistemas y a fin de divulgar el conocimiento. En el Apéndice 3. Descripción de esquemas XML. Representación del conocimiento, se describen los esquemas XML diseñados.

Capítulo 5

Sistema inteligente nítido de determinación taxonómica

La representación del conocimiento descrita en el capítulo anterior, nos permite confeccionar diferentes sistemas que utilicen los elementos detectados y categorizados, a raíz del estudio de las relaciones entre taxones y características. Como base para la evolución a sistemas inteligentes de determinación, se establece un sistema nítido de determinación taxonómica, capaz de utilizar los elementos nítidos del conocimiento e inferir una solución lo más adecuada posible a la descripción proporcionada por el usuario. Esta solución se llevará a cabo por medio de algoritmos simples que describiremos, de los cuales obtendremos no una única solución, si no el conjunto de posibles taxones que más se asemejan a las observaciones del usuario. Esta lista de taxones tendrá una ordenación priorizada, atendiendo al valor que hayan obtenido en la evaluación realizada (en función de los algoritmos empleados).

5.1. Selección de respuestas

El sistema presenta una lista de agrupaciones con sus correspondientes características, de tal manera que el usuario pueda evaluar cada una de ellas independientemente. El orden en el que se realiza esta evaluación no es prefijado por el sistema, es el propio usuario el que va escogiendo la siguiente característica a evaluar, pudiendo incluso rectificar juicios ya seleccionados en características previamente evaluadas. Los requisitos esquematizados que deben cumplir cualquiera de nuestros sistemas son los siguientes:

- **REQ1.** Elegir la siguiente característica a observar
 - **REQ1.1.** Seleccionar una respuesta tras evaluar la característica
 - **REQ1.2.** No seleccionar ninguna respuesta tras evaluar la característica, es posible que no conozca con certeza la respuesta.
- **REQ2.** Rectificar respuestas ya emitidas.
- **REQ3.** Eliminar una respuesta dada con anterioridad.

Si pensamos en los métodos dicotómicos clásicos los cuales obligan al usuario a seguir una consecución fija de observaciones cuya respuesta es obligatoria, o de lo contrario la determinación no puede continuar. Los sistemas de determinación que estudiaremos tanto nítidos como borrosos, cumplen los requisitos que acabamos de enumerar, presentando una ventaja clara sobre los métodos clásicos.

El modo en el que el sistema permite introducir o seleccionar las respuestas depende del tipo de característica que se quiera enjuiciar.

Características cuantitativas

Estas características medibles resultan las más sencillas a tratar por el sistema, ya que requiere únicamente que el usuario introduzca el valor medido numéricamente en el espécimen de estudio. El sistema para cumplir con los requisitos, debe permitir al usuario la no selección de respuestas, así como la eliminación de una respuesta deseleccionando la característica. En la Figura 5-1 podemos ver una posible implementación en un sistema de determinación.

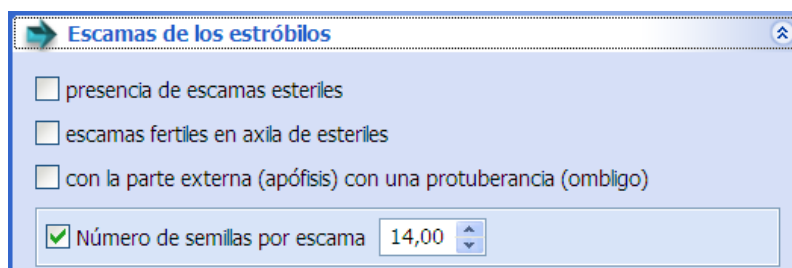


Figura 5-1. Ejemplo de selección de respuesta a una característica cualitativa.

Características cualitativas

Las características no medibles o cualitativas en los sistemas nítidos son tratadas como características booleanas en todos los casos, es decir, las características con atributo

aproximadas se comportan exactamente igual que las booleanas sin aportar información extra.

Estudiando el modo en el que el usuario puede emitir su respuesta, encontramos dos posibilidades:

- El usuario simplemente selecciona o marca positivamente una característica cuando es observada como tal en el espécimen. Si el usuario no es capaz de evaluar la característica o la evalúa negativamente, entonces esta aparecería deseleccionada. En la Tabla 5-1 podemos ver dos ejemplos de esta situación y su correspondiente representación en una posible implementación del sistema en la Figura 5-2.

Tabla 5-1. Características no unificadas.

Orden	"c" cuantitativa, "u" cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Grupo auto-excluyentes	Aproximadas	Orden relación proximadas	PINACEAE	Abies	Abies pinsapo	Abies alba	Pseudotsuga	Pseudotsuga menziesii
110			Características de hojas		4									
117			con dos bandas estomáticas en la cara inferior						1	3	3	3	0	0
120			cicatriz de la hoja circular						1	3	3	3	3	3
121			con canales resiníferos		5				1	3	3	3	3	3

Figura 5-2. Ejemplo de selección de características cualitativas.

- La segunda posibilidad es permitir al usuario para todas las características que pueda responder positiva o negativamente, si lo desea, a cada una de ellas. Es decir, el usuario tiene tres opciones, *sí*, *no*, *no sé*. Como vemos en el ejemplo de la Figura 5-3. La ventaja de esta opción es que es posible extraer información de las preguntas negativas y diferenciar cuando el usuario no conoce la respuesta frente a cuando la respuesta es negativa. Esta situación no es posible en la anterior aproximación, ya que la ausencia de marca puede significar ambas cosas.

no sé	sí	no	
<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	con dos bandas estomáticas en la cara inferior
<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	cicatriz de la hoja circular
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	con canales resiníferos

Figura 5-3. Ejemplo de selección *sí*, *no*, *no sé* de características cualitativas.

Si recordamos los posibles valores de las características cualitativas, estos podían ser (-1, 0, 1, 2, 3), la definición del valor -1 indicaba *No es posible. Ese taxón quedaría rechazado si se da la característica*. Por lo tanto, si tenemos un taxón que tiene valor -1 para la característica evaluada $R_{xy}(T_x, C_y) == -1$, sabemos que esa característica nunca podrá darse en ese taxón. Si permitimos al usuario introducir la respuesta *no*, podemos extraer información adicional para la evaluación de la regla de ese taxón, ya que la respuesta negativa concuerda con lo que sabemos de esa relación. La relación tiene valor -1 porque nunca se puede dar, y el usuario ha evaluado negativamente esa característica, por lo tanto se convierte en una premisa con información que puede ayudarnos a inferir la determinación con un grado superior de confianza al tener un mayor número de premisas evaluadas.

Características auto-excluyentes

Estos conjuntos se caracterizan en que no pueden ser observadas simultáneamente dos características del conjunto en un espécimen de estudio. La consecuencia lógica de esta definición es que en el momento en el que el usuario responda positivamente a una de las características del conjunto, se infiere automáticamente una respuesta negativa del resto de características del conjunto.

El modo en el que el usuario emite la respuesta vuelve a tener dos posibilidades como en el caso general anterior:

- El usuario únicamente puede seleccionar o deseleccionar una característica del conjunto. En el momento en el que realiza una selección, automáticamente el resto se convierten en no seleccionables, y su respuesta se infiere negativa. Por lo tanto si una característica del grupo está seleccionada, la no selección del resto (forzada por el sistema) implica una respuesta negativa y no desconocimiento como en el caso anterior. Sin embargo, la ausencia de selección positiva representa el desconocimiento sobre si alguna característica del conjunto es positiva, es decir, el usuario no es capaz de emitir un juicio positivo sobre

ninguno de los elementos del grupo de auto-exclusión. Podemos ver una posible implementación en la Figura 5-4

Figura 5-4 shows a dialog box titled "Ramificación" with a list of four options. The first option is "Ramificación monopódica", the second is "Ramificación simpódica", the third is "Ramificación verticilada regular" (which is selected with a green dot), and the fourth is "Ramificación verticilada irregular".

Figura 5-4. Ejemplo de un grupo de auto-exclusión, solo una posible selección.

- De la misma manera que describimos para características cualitativas, en los grupos de auto-exclusión también existe la posibilidad de que el usuario pueda formular su respuesta como *sí*, *no*, *no sé*. La casuística asociada sería la siguiente:
 - Si el usuario realiza una respuesta positiva sobre alguna de las características del grupo, entonces el resto de elementos del conjunto automáticamente reciben respuesta negativa.
 - Si el usuario no puede dar con certeza ninguna respuesta positiva puede ser por dos motivos. En primer lugar puede no conocer información de ninguna de las características, por lo que su respuesta final sería *no sé* para todas ellas. Existe una segunda posibilidad y es que aunque no pueda emitir respuesta positiva, si pueda apreciar una o varias respuestas negativas. En ningún caso una respuesta negativa infiere resultados para el resto de características. En el ejemplo de la Figura 5-5 podemos ver esta casuística representada en una posible implementación.

Figura 5-5 shows a dialog box titled "Ramificación" with three columns of radio buttons labeled "no sé", "sí", and "no". The first column has four radio buttons, the second has four, and the third has four. The first two columns have green dots next to the first two radio buttons in each row. The third column has a green dot next to the third radio button in the first row. The labels for the rows are "Ramificación monopódica", "Ramificación verticilada regular", "Ramificación simpódica", and "Ramificación verticilada irregular".

Figura 5-5. Ejemplo de un grupo de auto-exclusión con selección *sí*, *no*, *no sé*.

5.2. Inferencia nítida

Toda relación entre una característica y un taxón implica que dicha característica pasa a ser una premisa a evaluar en la regla de inferencia que describe al taxón involucrado. En la tabla de relaciones del capítulo anterior, vimos como estas relaciones tienen un valor asignado dependiendo del tipo de característica implicado, e incluso es posible marcar relaciones como desconocidas o ausencia de relación.

El procesamiento de la información contenida en el conocimiento, nos proporciona las premisas de inferencia nítida. Los distintos algoritmos (reglas) nítidos, utilizarán estas premisas para inferir el grado de determinación de cada taxón, de entre el conjunto de posibles taxones que componen la solución.

5.2.1. Premisas a evaluar y valor a asignar a dicha evaluación

La continua selección de respuestas por parte del usuario proporciona los valores de las premisas a evaluar en cada regla. A continuación, evaluaremos la interpretación de cada tipo de característica con sus posibles premisas y el tratamiento e inferencia de los valores. Estos valores, serán finalmente utilizados para el cálculo del grado de determinación de cada posible taxón.

Características cuantitativas

Tenemos una relación $R_{xy}(T_x, C_y)$ entre un taxón y una característica, donde C_y es una característica cuantitativa descrita por una lista de límites $[L_n]$, y la relación con el taxón T_x se compone de una lista de pares límite, valor $[(L_n, V_n)]$. Los intervalos que forman estos límites tendrán un grado de importancia a la hora de evaluar la premisa correspondiente en un sistema nítido. El grado asignado se lleva a cabo utilizando el baremo de las características cualitativas. Recordamos de nuevo la Tabla 4-3 en la nueva Tabla 5-2 para mayor facilidad.

Tabla 5-2. Conjunto completo de posibilidades asociadas a una relación $R_{xx}(T_x, C_x)$ para una característica cualitativa.

-1	No es posible. Ese taxón quedaría rechazado si se da la característica
0	No hay información
1	En ocasiones se da. Es posible, aunque no muy frecuente
2	Frecuentemente, normalmente.
3	Siempre se da.

El objetivo es definir una equivalencia entre los intervalos fijados por los límites de las relaciones y el baremo de posibilidad descrito. En primer lugar destacar que no todas las características ni todas las relaciones tienen todos los límites informados en el Conocimiento. Por lo tanto es necesario definir el valor de cada intervalo entre límites

suponiendo que pueden faltar algunos de ellos. Tenemos dos posibles grupo de límites (Min, MinInt, MaxInt, Max) y (Med, ES), estudiemos el primero de ellos. Si la característica y en concreto la relación con el taxón tiene todos los límites posibles, se establece como intervalo más probable el comprendido entre [MinInt, MaxInt], como la posibilidad más fuerte en el baremo es 3, asignamos este valor al intervalo. La premisa que hemos establecido es que si el resultado evaluado por el usuario reside en ese intervalo, tiene una posibilidad de grado 3 para ese taxón ya que es el intervalo más frecuente.

Si observamos el resto de intervalos, podemos ver que todos los valores que no estén comprendidos entre los límites inferirán un valor -1 para la premisa, dado que no corresponde a lo establecido por la relación. Los intervalos más alejados del núcleo central, tendrán como baremo 2, ya que ese trata de un caso frecuente pero no es el etiquetado como intervalo frecuente. La Figura 5-6 muestra esta situación. Si observamos el resto de imágenes, veremos cómo los intervalos se fijan en función de los límites de los que disponemos, y siempre podemos identificar un intervalo más frecuente que será el que tenga valor 3. En la Figura 5-7, podemos ver la equivalencia en el grupo (ES, med).

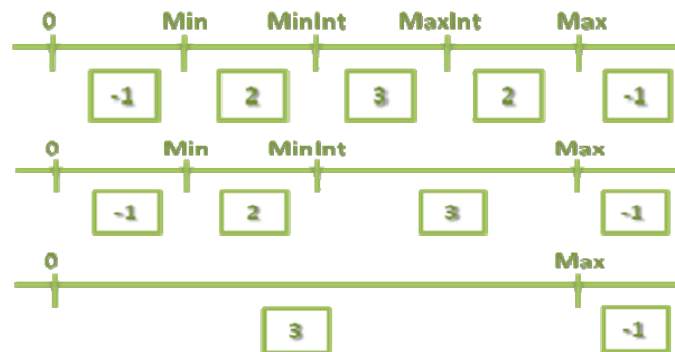


Figura 5-6. Valores inferidos en intervalos delimitados por límites (Min, MinInt, MaxInt, Max) para una relación $R_{xx}(T_x, C_x)$. Tres ejemplos (en este orden) con todos los límites, sin un límite, solo con un límite.

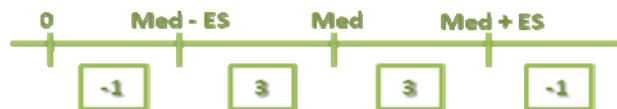


Figura 5-7. Valores inferidos en intervalos delimitados por límites (Med, ES).

Características cualitativas

Si una característica cualitativa tiene relación positiva con un taxón determinado, esta premisa se convierte en elemento de la regla de inferencia del taxón. Los posibles valores positivos (ver Tabla 5-2) son 1, 2 y 3. Estos valores serán los que utilice el

algoritmo escogido para obtener el grado de determinación. Si la relación tiene valor 0 o no tiene valor, la premisa no participa en la regla, ya que no se dispone de información. El valor -1 representa una premisa excluyente, el modo de utilizar esta premisa ya ha sido comentado con anterioridad; si el usuario selecciona una característica con valor -1 en la relación con un taxón, el taxón se elimina automáticamente del conjunto solución.

Esta inferencia que acabamos de describir se realiza con características que han sido seleccionadas con respuestas positivas. Como vimos en el punto 5.1, existe una primera posibilidad en la cual los algoritmos únicamente utilizan como premisas características con respuesta positiva, el usuario o responde positivamente o no responde nada, no tiene la opción de respuesta negativa.

La relación $R_{ij}(T_i, C_j)$ se evalúa:

- $V_{ij} == -1 \rightarrow$ Premisa excluyente, el taxón se elimina automáticamente de la solución.
- $V_{ij} == 0 \rightarrow$ Sin información. La característica no es una premisa.
- $V_{ij} == 1, 2 \text{ o } 3 \rightarrow$ Premisa positiva. El valor será el utilizado en la evaluación de la premisa en las reglas.

Si contemplamos la segunda posibilidad en la que el usuario tiene la opción de responder *sí*, *no*, *no sé* además de la información que acabamos de concluir para respuestas positivas, tenemos nueva información que podemos modelar. Si el usuario ha respondido negativamente a la característica C_j (no se da la característica C_j en el espécimen), para toda relación $R_{ij}(T_i, C_j)$ entre cualquier taxón T_i y la característica C_j ; si el valor de dicha relación es -1, la característica se convierte en premisa positiva, ya que para ese taxón esa característica nunca se da (concuerda con lo observado por el usuario). Por el contrario si la relación se evalúa con valor positivo 3, pasará a ser premisa excluyente evaluada con valor -1, ya que para ese taxón la característica se tiene que dar siempre y el usuario ha emitido una respuesta negativa que indica que no se da en el espécimen de estudio.

La relación $R_{ij}(T_i, C_j)$ se evalúa:

- $V_{ij} == -1 \rightarrow$ Premisa que se convierte en positiva. Nunca se da la característica, el usuario la observa negativa. Se evalúa en la regla con el grado máximo 3.
- $V_{ij} == 0 \rightarrow$ Sin información. La característica no es una premisa.

- $V_{ij} == 1 \rightarrow$ Premisa positiva. No es frecuente que se dé, por lo tanto si el usuario ha respondido negativamente, se está cumpliendo la premisa en un grado alto (aunque menor que el anterior). La evaluación en la regla se realiza con grado 2.
- $V_{ij} == 2 \rightarrow$ Premisa positiva. Frecuentemente se da, pero es posible que no se dé, por lo tanto se está cumpliendo la premisa pero con un grado mucho menor. Se evalúa la regla con grado 1.
- $V_{ij} == 3 \rightarrow$ Premisa excluyente. Se tiene que dar siempre, el usuario ha emitido una respuesta negativa. Se evalúa la regla con grado -1. El taxón se elimina automáticamente de la solución.

Características cualitativas auto-excluyentes

El comportamiento de la primera posibilidad es equivalente en las respuestas auto-excluyentes, aquellas características que hayan sido seleccionadas son las que participan como premisas en las reglas. El valor de la relación se convierte en la evaluación de la premisa tal y como hemos visto.

La segunda posibilidad no es aplicable directamente a características auto-excluyentes, veamos un ejemplo para ilustrar el problema en la Tabla 5-3. En la primera respuesta el usuario nos indica que observa la característica C_1 , la relación con el taxón tiene valor máximo 3, por lo que se evaluará una premisa con valor 3. En la segunda respuesta, si aplicamos directamente las reglas vistas anteriormente, al tener un “no” en dos premisas que el taxón tiene valoradas con -1, tendríamos que evaluar la premisa con un 3, grado máximo. Es decir, que estaríamos obteniendo dos premisas con grado máximo 3. Si comparamos ambos casos, vemos que la inferencia en el segundo caso va a ser siempre mayor que la del primer caso y es un error. La respuesta 2 nos está proporcionando una posibilidad alta de que se dé C_1 pero siempre menor que la respuesta 1, ya que en esta respuesta es seguro que se da C_1 .

Tabla 5-3. Ejemplo baremos características auto-excluyentes con respuestas negativas.

	T_i	Respuesta 1	Respuesta 2
C_1	3	si \Rightarrow 3	-
C_2	-1	-	-
C_3	-1	-	no \Rightarrow 3
C_4	-1	-	no \Rightarrow 3

1 premisa valor 3	2 premisas valor 3	Evaluación
----------------------	-----------------------	------------

Sin embargo, la información que nos proporciona emitir una respuesta negativa puede ser aprovechada realizando un modelo diferente. Tenemos un conjunto de autoexclusión, si existe un número mayor que 0 de características con respuesta negativa, cuyo valor en la relación con el taxón es para todas -1, podemos inferir cierto grado de posibilidad para que se cumpla la característica C_1 . Obtenemos una premisa con un único valor, cuya evaluación se realiza en función del porcentaje de variables auto-excluyentes, con valor -1 en la relación, respondidas negativamente del mismo grupo al que pertenece. Hay muchas posibilidades para baremar estos porcentajes como evaluaciones de la premisa. En un posterior estudio, se podrían implementar algunas de estas posibilidades y estudiar los resultados obtenidos para determinar la más adecuada.

5.2.2. Ejecución de reglas

Si tenemos un sistema con cinco características de las cuales el usuario ha evaluado 2 positivamente y 3 desconocidas, tenemos la siguiente información contenida en la Tabla 5-4, donde el valor 1 indica selección positiva y el valor 0 desconocido. Por otro lado, si nuestro conocimiento contiene tres taxones, con las consecuentes relaciones con nuestras cinco características (ver Tabla 5-5), conocemos el valor de estas relaciones y podemos reunir todas las premisas de la regla junto con su valor evaluado.

Tabla 5-4. Ejemplo conjunto de características seleccionadas.

	C_1	C_2	C_3	C_4	C_5
Selección positiva	S_1	S_2	S_3	S_4	S_5
Ejemplo	1	0	1	0	0

Tabla 5-5. Matriz de valores para las relaciones.

R_{ij}	C_1	C_2	C_3	C_4	C_5
T_1	V_{11}	V_{12}	V_{13}	V_{14}	V_{15}
T_2	V_{21}	V_{22}	V_{23}	V_{24}	V_{25}
T_3	V_{31}	V_{32}	V_{33}	V_{34}	V_{35}

Para un taxón T_i su regla de inferencia tendrá como entrada el conjunto de sus premisas evaluadas seleccionadas $[(C_1 * S_1), (C_2 * S_2), (C_3 * S_3), (C_4 * S_4), (C_5 * S_5)]$. En el ejemplo, el conjunto de premisas sería (C_1, C_3) con valores $[V_{i1}, V_{i3}]$. Sobre estos conjuntos aplicamos los algoritmos y obtenemos el valor inferido de determinación para el taxón.

Para la segunda posibilidad planteada, la información proporcionada por el usuario sería por ejemplo la que ilustra la Tabla 5-6, como podemos ver para la característica C_4 estamos obteniendo información extra respecto al ejemplo de la Tabla 5-4, en el

ejemplo anterior no existía información diferenciada entre las características 2,4 y 5, mientras que en este caso se observa un resultado negativo para C₄. El grado de inferencia de la solución se ve aumentado con esta nueva premisa. Para poder profundizar completamente en esta solución sería necesario resolver el problema de las características auto-excluyentes visto anteriormente. Aunque la estimación en porcentajes aporta mucha información, se pospone para un estudio posterior a fin de centrar el que ahora nos ocupa en la creación de sistemas de determinación taxonómica borrosos.

Tabla 5-6. Ejemplo conjunto de características con respuestas sí, no, ni sé.

	C ₁	C ₂	C ₃	C ₄	C ₅
Selección positiva	S ₁	S ₂	S ₃	S ₄	S ₅
Ejemplo	sí	no sé	si	no	no sé

5.2.3. Algoritmos

Las reglas aplicadas en el sistema nítido no son más que algoritmos sencillos que interpretan el cálculo del grado de inferencia en la determinación, partiendo de las premisas evaluadas que vimos anteriormente.

Todo algoritmo nítido tiene como norma eliminar un taxón si alguna de sus premisas es evaluada como -1. Tras realizar este primer paso, se procede a calcular el grado de inferencia en función de las premisas evaluadas positivas (valores 1, 2 y/o 3).

- **Algoritmo Suma Valores**

Este algoritmo primeramente elimina aquellos taxones que contienen entre sus premisas alguna con evaluación -1. Procede entonces a sumar todas las premisas evaluadas con valores comprendidos en (1, 2, 3), tanto para características cualitativas, como para cuantitativas. Como vimos en el punto anterior, una característica cuantitativa se barema con los mismos grados que una cualitativa, atendiendo al intervalo en el que está comprendido el valor numérico introducido por el usuario al observar dicha característica.

La lista ordenada obtenida, se ordena por suma de valores máxima, es decir el que haya obtenido el mayor número se considera el primero de la lista de soluciones. Esta lista es el conjunto solución inferido

- **Algoritmo “Medallero”**

Este algoritmo, de la misma manera que el anterior, elimina en primera instancia aquellos taxones que contienen entre sus premisas alguna con evaluación -1. La

entrada del algoritmo es una lista de evaluaciones para cada premisa que toman valores 1, 2 y/o 3. En este caso en vez de proceder a la suma de estos baremos, el algoritmo cuenta para cada taxón, el número de premisas evaluadas con máximo grado 3, el número de grado 2 y el de grado 1. Una vez dispone de las tres cantidades para cada taxón no eliminado, los ordena primero por el mayor número de evaluaciones de grado 3. A igualdad de cantidad se procede a ordenar por el mayor número de grado 2 y finalmente si es necesario por el grado 1.

La lista ordenada obtenida (número de treses, número de doses, número de unos), es el conjunto solución inferido, ordenado por el que ha obtenido un grado de determinación mayor.

Podemos ver en la Figura 5-8 una posible implementación del sistema nítido con las dos posibles soluciones derivadas del uso de ambos algoritmos.

Figura 5-8. Ejemplo de visualización en una posible implementación.

Capítulo 6

Sistema inteligente borroso de determinación taxonómica

En el capítulo anterior estudiamos las bases de un sistema nítido que utiliza la base de conocimiento diseñada en el Capítulo 4. En el estudio del sistema se han identificado los distintos tipos de características como premisas, el modo de evaluación de estas premisas se ha modelado en función de los valores hallados en las relaciones entre taxones y características. Así mismo se han diseñado reglas para cada posible taxón del conocimiento que, gracias a la evaluación de las premisas, son capaces de inferir un grado de determinación taxonómica para cada taxón. El conjunto de taxones ordenados con su correspondiente baremo, forman el conjunto solución. Este conjunto permite al usuario evaluar la mejor solución, en vez de que sea el propio sistema el que proporcione una solución única. Si bien siempre podemos decir que la solución con mayor grado es la más óptima de las evaluadas por el sistema.

Si recapitulamos sobre lo aprendido en el sistema nítido, tenemos premisas, reglas, operadores e inferencia, es decir, tenemos identificados los elementos bases de un sistema borroso. La primera labor que llevaremos a cabo en este capítulo, es proporcionar herramientas extras que flexibilicen las posibles respuestas que puede emitir el usuario, utilizando razonamiento aproximado. Identificaremos el conjunto de características con sus etiquetas lingüísticas modeladas mediante conjuntos borrosos, diseñaremos sus funciones de pertenencia (utilizando la herramienta Xfuzzy para su modelado) y finalmente realizaremos la inferencia borrosa por medio de las reglas construidas.

Los requisitos definidos para sistemas nítidos se aplicarán de la misma manera, el sistema deberá permitir al usuario:

- **REQ1.** Elegir la siguiente característica a observar
 - **REQ1.1.** Seleccionar una respuesta tras evaluar la característica
 - **REQ1.2.** No seleccionar ninguna respuesta tras evaluar la característica, es posible que no conozca con certeza la respuesta.
- **REQ2.** Rectificar respuestas ya emitidas.
- **REQ3.** Eliminar una respuesta dada con anterioridad.

6.1. Características

6.1.1. Características Cuantitativas

Las características medibles numéricamente, llamadas cuantitativas, están determinadas por unos límites establecidos, como vimos en la representación del conocimiento (recuperamos la representación de los mismos en la Tabla 6-1). El modo de selección del resultado, estudiado en los sistemas nítidos, será el mismo que proporcionaremos en el sistema borroso al usuario; el cual permite al usuario introducir un valor numérico y seleccionar o deseleccionar la característica si desea evaluar o eliminar su respuesta respectivamente.

Tabla 6-1. Representación gráfica de las relaciones para características cuantitativas.

		T ₁	T ₂	T ₃	T ₄
C ₁	L ₁	R ₁₁	R ₁₂	R ₁₃	R ₁₄
	L ₂	R ₂₁	R ₂₂	R ₂₃	R ₂₄
C ₂	L ₃	R ₃₁	R ₃₂	R ₃₃	R ₃₄
	L ₄	R ₄₁	R ₄₂	R ₄₃	R ₄₄
	L ₅	R ₅₁	R ₅₂	R ₅₃	R ₅₄

Toda relación existente entre un taxón y una característica numérica, está compuesta por los valores que dichos límites toman en la definición del taxón. A partir de cada una de estas características y los intervalos que describen sus límites, se modela un conjunto borroso trapezoidal para cada relación. La ventaja principal que ofrece al usuario, frente a un sistema nítido, es que no estamos eliminando directamente un

taxón como posible solución si la respuesta del usuario es cercana al intervalo descrito para ese taxón. A continuación veremos dos posibles variantes.

6.1.1.1. Intervalo máximo

Los valores cercanos al intervalo de máximo grado de pertenencia, sufren una bajada en dicho grado paulatinamente. De esta manera la premisa evaluada no tendrá grado de pertenencia máximo pero tampoco será descartada, ofreciendo información adicional que puede utilizarse en la regla de inferencia de ese taxón. Del mismo modo que hemos bajado el grado de pertenencia de la premisa, el grado inferido, mediante la regla de dicho taxón, también se verá afectado y pasará a tener un grado “un poco menor” de pertenencia a dicha especie.

El resultado obtenido tras aplicar este comportamiento es un efecto de computación suave.

Hemos visto que para cada relación (taxón, característica) tendremos una serie de límites informados. Cabe destacar que existen casos en los que la característica tiene un número mayor de posibles límites mientras que, los límites que realmente utilizan cada una de las relaciones con esa característica y los diferentes taxones, son un subconjunto de ellos. Por lo tanto el sistema de generación de conjuntos borrosos tendrá que poder adaptarse a cada caso particular.

Como sabemos tenemos dos grupos de límites (Min, MinInt, MaxInt, Max) y (Med, ES), los cuales pueden estar presentes de cada grupo, desde 1 a todos los elementos del grupo. En cualquier caso se escogerá siempre el intervalo máximo, y no se diferenciará dentro de él subintervalos, si por ejemplo tenemos una relación con los límites (Min, MinInt, MaxInt, Max), el intervalo máximo que tendrá el mayor grado de pertenencia (valor 1), será el comprendido entre Min y Max, y no se diferenciarán los intervalos resultantes de disponer de MinInt y MaxInt. Si bien en el siguiente apartado vemos una forma de obtener más información de estos subintervalos.

A continuación veremos varios ejemplos que describen el modo en el que el sistema autogenera los trapecios en función de los límites disponibles. Cada uno de los conjuntos borrosos están descritos por una función de pertenencia trapezoidal con cuatro puntos (a, b, c, d) que definen el trapecio, ver Figura 6-1 con trapecio genérico. Dependiendo de los límites involucrados en la relación, los puntos tomarán los valores de dichos límites. Las pendientes descendientes a cada lado del intervalo de grado máximo, vendrán descritas por los puntos a y d. Estos puntos a y d se calcularán a una distancia, de los puntos b y c respectivamente, medida en función de la anchura del intervalo. Si la anchura entre b y c es x, la distancia aplicada será del 10% de x, este

porcentaje se implementará variable para poder ajustarlo en futuros estudios. Podría incluso no ser un porcentaje fijo para todos los casos si no inferido o aprendido.

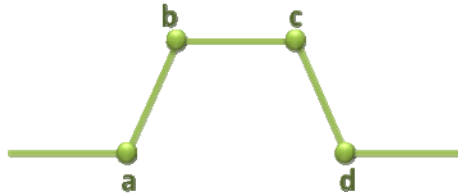


Figura 6-1. Función de pertenencia trapezoidal genérica.

Cada característica tiene una etiqueta lingüística por cada taxón relacionado. De tal manera que obtenemos tantos conjuntos borrosos, y sus funciones de pertenencia correspondientes, como taxones existen relacionados con la característica.

- **Ejemplo 1.** Altura (m). Tabla 6-2.

Esta característica tiene un único límite **max**, por lo tanto tomaremos el mínimo como 0 y el margen suave superior calculado con el 10% sobre el valor del límite max en cada relación (taxón, característica). Es decir, el 10% de la diferencia del intervalo $\rightarrow (\text{max} - \text{min}) * 0.1$. Aplicando estas reglas de generación sobre dos taxones de ejemplo obtenemos los límites descritos en la Tabla 6-3.

Tabla 6-2. Ejemplo 1. Altura (m)

Orden	"c" cuantitativa, "u" cualitativa	Agrupación	Características	Tipo dato cuantitativo	Abies	Pinus
14	c		Altura (m)	Max	50	40
15						

Tabla 6-3. Puntos que describen el trapecio de cada una de las funciones de pertenencia de los conjuntos borrosos de los taxones Abies y Pinus.

	a	b	c	d	valor intervalo
Abies	min - 10%	min	max	max + 10%	max - min
	0	0	50	55	50
Pinus	min - 10%	min	max	max + 10%	max - min
	0	0	40	44	40

A continuación vemos la gráfica resultante modelada en Xfuzzy para la etiqueta lingüística *mf_Abies*. Aparece marcada la función correspondiente en color rojo y los límites establecidos en la parte izquierda.

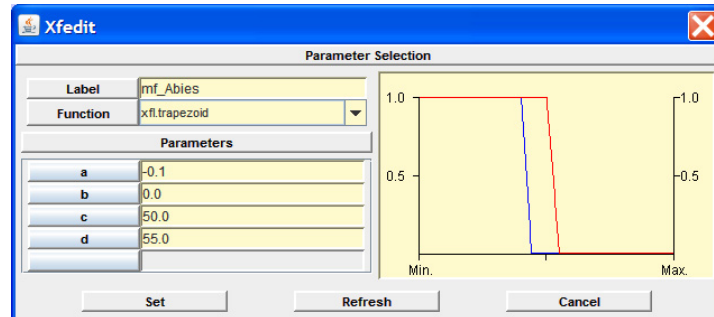


Figura 6-2. Ejemplo conjunto borroso *mf_Abies* de la característica *t_Altura_m*.

- **Ejemplo 2.** Longitud de los estróbilos (cm)

El siguiente ejemplo presenta la casuística anteriormente comentada, en la cual dos relaciones de dos taxones diferentes con una misma característica, pueden tener diferentes límites evaluados. Además observaremos que una de las funciones obtenidas es una función trapezoidal clásica.

Tabla 6-4. Ejemplo 2. Longitud de los estróbilos (cm)

Orden	"c" cuantitativa, "c" cualitativa	Agrupación	Características	Tipo dato cuantitativo	PINACEAE	Abies
169	c		Longitud de los estróbilos (cm)			
170				Max	22	20
171				Min		10

Tabla 6-5. Puntos que describen el trapecio de cada una de las funciones de pertenencia de los conjuntos borrosos de los taxones PINACEAE y *Abies*.

	a	b	c	d	valor intervalo
PINACEAE	min - 10%	min	max	max + 10%	max - min
	0	0	22	24,2	22
Abies	min - 10%	min	max	max + 10%	max - min
	9	10	20	21	10

A continuación, Figura 6-3, vemos la gráfica resultante modelada en Xfuzzy para la etiqueta lingüística *mf_Abies*. La función forma un trapecio clásico, ya que tiene valores para los cuatro puntos. Sin embargo la gráfica para PINACEAE, Figura 6-4, es diferente

ya que no tiene uno de los límites disponibles en la característica, su valor mínimo comienza en el 0.

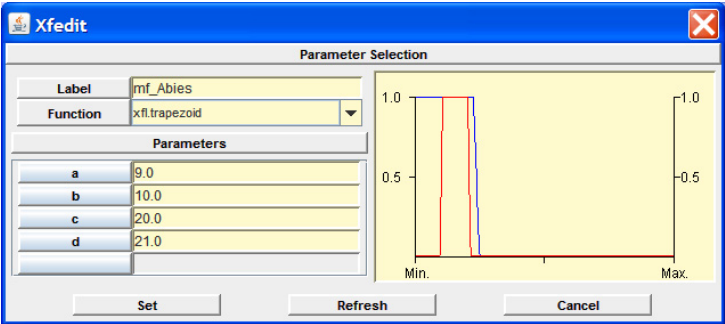


Figura 6-3. Ejemplo conjunto borroso *mf_Abies* de la característica *t_Longitud_estróbilos_cm*.

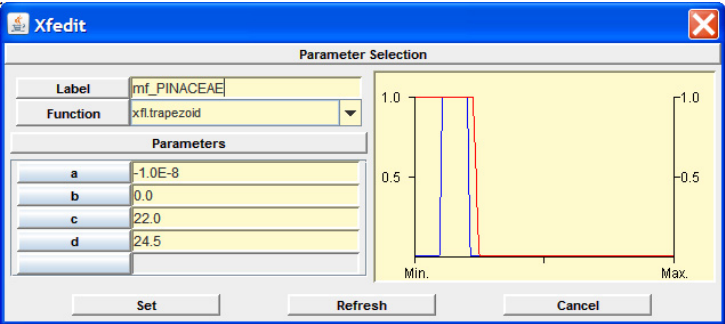


Figura 6-4. Ejemplo conjunto borroso *mf_PINACEAE* de la característica *t_Longitud_estróbilos_cm*.

• **Ejemplo 3.** Longitud de las hojas (mm)

El siguiente ejemplo presenta la casuística anteriormente comentada, en la cual dos relaciones de dos taxones diferentes con una misma característica, pueden tener diferentes límites evaluados. Además observaremos que una de las funciones obtenidas es una función trapezoidal clásica.

Tabla 6-6. Ejemplo 3. Longitud de las hojas (mm)

Orden	"c" cuantitativa, " " cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Grupo auto-excluyentes	Aproximadas	Orden relación proximadas	Abies	Pinus halepensis
103	c		Longitud de las hojas (mm)							
104				Max					30	150
105				Min					6	60
106				MaxInt						120

Tabla 6-7. Puntos que describen el trapecio de cada una de las funciones de pertenencia de los conjuntos borrosos de los taxones *Abies* y *Pinus halepensis*.

	a	b	c	d	valor intervalo máximo
Abies	min – 10%	min	max	max + 10%	max - min
	3,6	6	30	32,4	24
Pinus halepensis	min – 10%	minInt	max	max + 10%	max - minInt
	51	60	150	159	90

A continuación, Figura 6-5 y Figura 6-6, vemos la gráfica resultante modelada en Xfuzzy para los dos conjuntos borrosos diseñados. Las pendientes de ambos trapecios, son calculadas con el 10% del intervalo máximo.

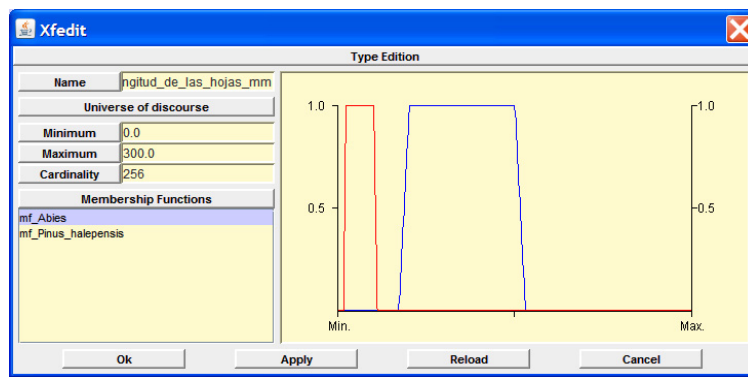


Figura 6-5. Intervalo máximo. Ejemplo conjunto borroso *mf_Abies* de la característica *t_Longitud_de_las_hojas_mm*.

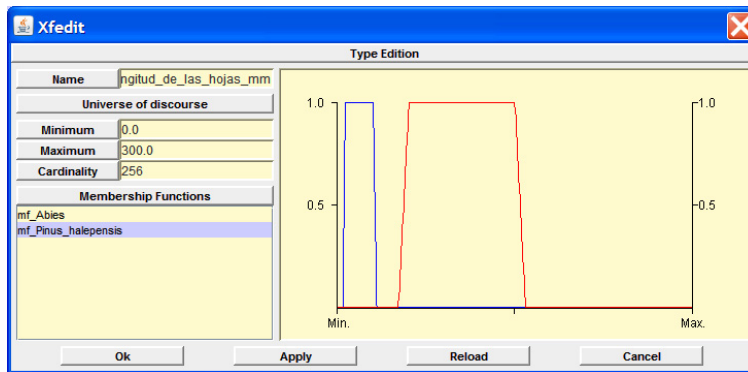


Figura 6-6. Intervalo máximo. Ejemplo conjunto borroso *mf_Pinus_halepensis* de la característica *t_Longitud_de_las_hojas_mm*.

6.1.1.2. Intervalo frecuente

En el caso anterior hemos contemplado como intervalo de grado 1 el intervalo máximo. En este apartado estudiamos cómo conseguir información extra de aquellas características que tengan varios subintervalos dentro del intervalo máximo. En el

conjunto de límites (Min, MinInt, MaxInt, Max) estudiado en punto 4.2.2, encontrábamos las definiciones de la Tabla 6-8.

Tabla 6-8. Descripción de límites de características cuantitativas.

Max	Máximo
Min	Mínimo
MaxInt	Máximo del intervalo frecuente
MinInt	Mínimo del intervalo frecuente
Med	Media
ES	Error estándar

Cuando para una relación entre un taxón y una característica tenemos todos los límites de este grupo, encontramos que [MinInt, MaxInt] representan el intervalo frecuente de la relación. Es decir, es el intervalo de mayor grado de probabilidad, si recordamos el sistema nítido, a este intervalo le asignábamos el valor 3 como grado de frecuencia máxima. Los intervalos descritos [Min, MinInt] y [MaxInt, Max], tienen una alta probabilidad pero no son tan frecuentes como el intervalo frecuente. Si recordamos los sistemas nítidos, a estos intervalos les asignaba el valor 2. Gracias a las funciones trapezoidales podemos conseguir una representación más suave de esta situación, en vez de asignar el valor 2 a todo el intervalo, podemos modelar la función de tal manera que los valores más cercanos al intervalo frecuente tengan una graduación mayor y a medida que nos alejamos disminuya suavemente el grado.

Los puntos b y c de la Figura 6-7 delimitan el intervalo frecuente [MinInt, MaxInt], el punto a corresponde al Min y el punto d al Max

Adicionalmente, igual que hicimos con el intervalo máximo, los conjuntos borrosos nos permiten no eliminar automáticamente una premisa, si el valor que el usuario ha introducido es realmente cercano a los límites que definen la característica. De esta manera en el apartado anterior definimos un porcentaje (10% variable) que delimitaba la pendiente del trapecio. Este mismo criterio se puede aplicar al nuevo modelo, de tal manera que el punto a se calcule sobre el extremo Min menos un épsilon calculado a partir del porcentaje de la diferencia del intervalo frecuente:

- $a \rightarrow \text{Min} - \text{épsilon}$, donde épsilon es el 10% de la diferencia (c-b)
- $b \rightarrow \text{MinInt}$
- $c \rightarrow \text{MaxInt}$
- $d \rightarrow \text{Max} + \text{épsilon}$, donde épsilon es el 10% de la diferencia (c-b)

En la Figura 6-7, se perciben dos nuevos puntos x e y , estos puntos representan los valores de Min y Max respectivamente.

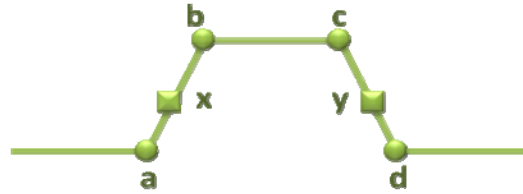


Figura 6-7. Función de pertenencia trapezoidal genérica.

Los dos primeros ejemplos del apartado anterior tendrían la misma equivalencia en esta opción, ya que únicamente disponen de dos límites. Sin embargo el ejemplo número 3, ilustra la diferencia entre ambas casuísticas, vemos a continuación el resultado del intervalo frecuente.

La Tabla 6-9 muestra los límites obtenidos con el intervalo frecuente. A continuación en Figura 6-8, vemos los conjuntos borrosos trapezoidales obtenidos.

Tabla 6-9. Intervalo frecuente. Puntos que describen el trapecio de cada una de las funciones de pertenencia de los conjuntos borrosos de los taxones *Abies* y *Pinus halepensis*.

	a	x	b	c	y	d	intervalo frecuente
Abies	min - 10% 3,6	N/A	min 6	max 30	N/A	max + 10% 32,4	max - min 24
Pinus halepensis	min - 10% 57	min 60	minInt 120	max 150	N/A	max + 10% 153	max - minInt 30

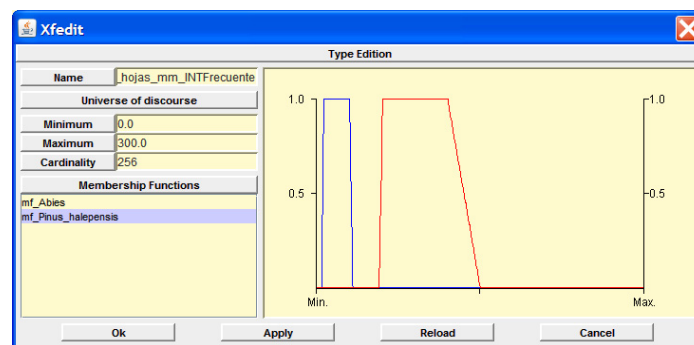


Figura 6-8. Intervalo frecuente. Ejemplo conjunto borroso $mf_Pinus_halepensis$ de la característica $t_Longitud_de_las_hojas_mm$.

6.1.1.3. Etiquetas lingüísticas

Como hemos visto en los puntos anteriores, cada característica cuantitativa descrita por el conjunto de sus límites, tendrá un número de etiquetas lingüísticas igual al número total de taxones para los cuales existe una relación entre la característica y el taxón.

6.1.2. Características Cualitativas

Cada tipo de característica detectado en el estudio del conocimiento tiene un tratamiento diferente. Procederemos a explicar cada casuística por separado.

6.1.2.1. Características cualitativas booleanas

Estas características son aquellas que únicamente tiene como posible respuesta sí y no, las funciones de pertenencia descritas son planas, y las premisas toman grados de pertenencia 0 o 1.

Tenemos entonces dos conjuntos borrosos, uno para la etiqueta lingüística no y con grado 1 de pertenencia en el valor 0 del eje, y grado 0 de pertenencia en el valor 1 del eje. El segundo conjunto corresponde a la etiqueta sí con grado de pertenencia 0 en el valor 0 del eje y grado 1 en el valor 1 de eje. La representación de los conjuntos borrosos genéricos se pueden observar en la Figura 6-9.

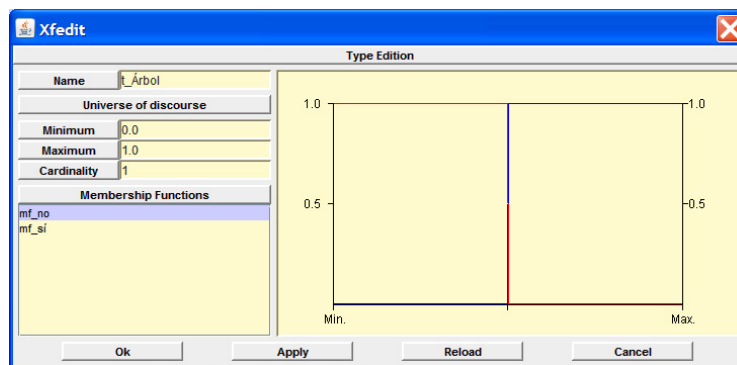


Figura 6-9. Ejemplo conjuntos borrosos *mf_no* y *mf_si* de la característica booleana *t_Árbol*.

Si observamos el ejemplo de la Tabla 6-10 vemos una característica booleana Árbol (fanerófito), y dos taxones con valores para la relación con dicha característica. Nos preguntamos entonces, cómo definir el conjunto borroso para cada una de estas relaciones. Recordemos que los valores 2 y 3 de la tabla indican los grados de mayor frecuencia de aparición de la característica para el taxón de la relación. Por el contrario, el valor -1 y 1 indican la menor frecuencia de aparición, más aún el valor -1 indica la imposibilidad de aparición. Por lo tanto:

- Asignaremos a todas aquellas relaciones con valores 2 o 3 el conjunto borroso sí.
- Asignamos a las relaciones con valores -1 y 1 el conjunto borroso no.

Tabla 6-10. Ejemplo de característica booleana

Orden	"C" cuantitativa, "c" cualitativa	Agrupación	Características	TAXACEAE	EPHEDRACEAE
10		Biotipo			
11			Árbol (fanerófito)	3	-1
12			Arbusto (nanofanerófito)	-1	2
13			Matorral bajo (caméfito)	-1	1

Como vimos en los sistemas nítidos, si el sistema permite respuestas *no*, *sí*, *no sé*, es complicado modelar la situación de las características auto-excluyentes. Hemos dejado esta evolución para un estudio posterior, por lo tanto necesitamos ver la equivalencia de las características booleanas, en un sistema que únicamente permite seleccionar positivamente una respuesta o deseleccionarla, esta última opción puede indicar desconocimiento o respuesta negativa indistintamente y no tenemos modo de distinguirlo.

Si pensamos entonces en estas dos únicas opciones, la selección positiva hará que automáticamente las relaciones con valores 2 o 3 pasen a ser premisas de las reglas de inferencia con grado de pertenencia 1. Este grado es tomado directamente de la función de pertenencia $mf_{sí}$ descrita anteriormente. Aunque no podamos tener respuesta negativa, la información positiva es muy útil, ya que todas las relaciones con valores -1 y 1 para la característica, serán premisas para las reglas de los taxones involucrados en dichas relaciones, y el grado de pertenencia será 0 según la función mf_{no} .

De esta manera estamos defuzzificando al pasar de información borrosa a nítida, es decir de conjuntos borrosos a conjuntos nítidos. La defuzzificación más habitual es escoger el centro de gravedad -de un fuzzy set continuo- y asociarle un número. En este caso la defuzzificación ha consistido en asociar los valores -1 y 1 al *no* y los valores 2 y 3 al *sí*. Se podrían estudiar otras defuzzificaciones y evaluar sus resultados para escoger la más adecuada. Incluso sería posible preguntar a los expertos para ajustar el modo.

Como hemos visto, las características booleanas solo pueden ser evaluadas con respuesta positiva (marcadas), o no evaluadas (no marcadas) que implica indistinguiblemente respuesta desconocida o negativa. Si obtenemos una respuesta

positiva del usuario, todas las relaciones con valor 2 o 3 serán una premisa evaluada con grado de pertenencia máximo 1 en la regla. Además, todas las relaciones con valores -1 o 1, son automáticamente premisas con grado cero, estas premisas interfieren en las reglas de inferencia por lo que el taxón queda automáticamente descartado. Para modelar esta situación, vemos que tenemos un conjunto borroso que indica que la característica ha sido observada en el espécimen *mf_si*, podemos definir un segundo conjunto borroso que indique que la característica no ha sido observada *mf_no* (simétrico respecto a la recta $y = 1/2$).

En esta defuzzificación hemos perdido parcialmente la información que los valores 1 y 2 aportaban al sistema. A continuación veremos un modelo diferente, compuesto por cuatro conjuntos borrosos que utilizan dicha información.

Alternativa

Si pensamos en el valor 1 de la relación $R_{ij}(T_i, C_j)$, el caso más frecuente de este valor es el conjunto borroso *no*, ya que indica que es poco frecuente que se dé la característica C_j en el taxón T_i . Sin embargo en la defuzzificación que hemos realizado, estaríamos descartando este taxón cuando la respuesta a la característica C_j fuera positiva, pese a que la propia definición de este valor nos indica que aunque no es muy frecuente, a veces se da.

Esta misma situación ocurre con el valor 2, al que hemos asignado el conjunto borroso *sí*. Si se infiere una respuesta negativa (ejemplo auto-excluyentes), estaríamos eliminando directamente el taxón (marcado con valor 2 en la relación con esa característica) como solución, pese a que la propia definición del valor nos indica que es bastante frecuente que se dé, pero que puede no darse.

Para suavizar estos límites establecidos, podemos crear dos nuevos fuzzy sets, de tal manera que:

- el fuzzy set *no* cambia su etiqueta lingüística por *extermadamente_no* y se asocia a relaciones con valores -1,
- el fuzzy set *si* cambia su etiqueta lingüística por *extermadamente_si* y se asocia a relaciones con valores 3,
- se crea un nuevo fuzzy set con la etiqueta *no*, la función de pertenencia se suaviza, no baja bruscamente a grado 0 en el límite con el valor 1, si no que se degrada paulatinamente, incluyendo “un poco” de *sí*.

- se crea un nuevo fuzzy set con la etiqueta *sí*, la función de pertenencia se suaviza, no baja bruscamente a grado 0 en el límite con el valor 0, si no que se degrada paulatinamente, incluyendo “un poco” de *no*.

Podemos ver gráficamente la solución en la Figura 6-10 y Figura 6-13. La cuestión que tenemos que abordar en este punto es asignar un valor a la respuesta *no*, y otro a la respuesta *sí*, ya que si asignamos $no \rightarrow 0$ y $si \rightarrow 1$, estamos perdiendo igualmente la información de los conjuntos borrosos suaves que acabamos de crear. El valor a asignar debe ser variable y ajustable, para que en futuros experimentos se puedan encontrar resultados que indique el valor que más se ajusta o que mejores resultados da. En esta investigación tomaremos los valores medios entre los extremos Min y Max y $\frac{1}{2}$ del eje X. Es decir, valor 0.25 para el *no* y 0.75 para el *sí*.

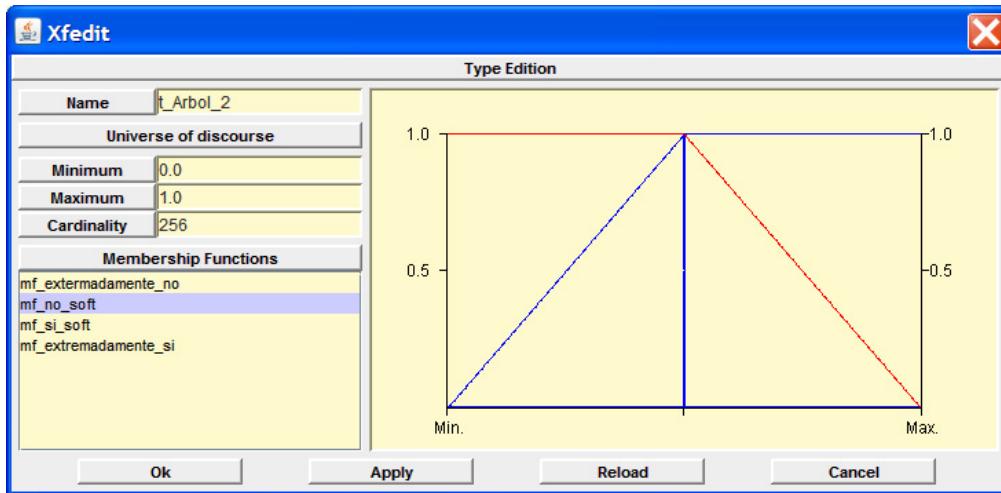


Figura 6-10. Ejemplo conjuntos borrosos *mf_no_soft* de la característica booleana *t_Arbol*.

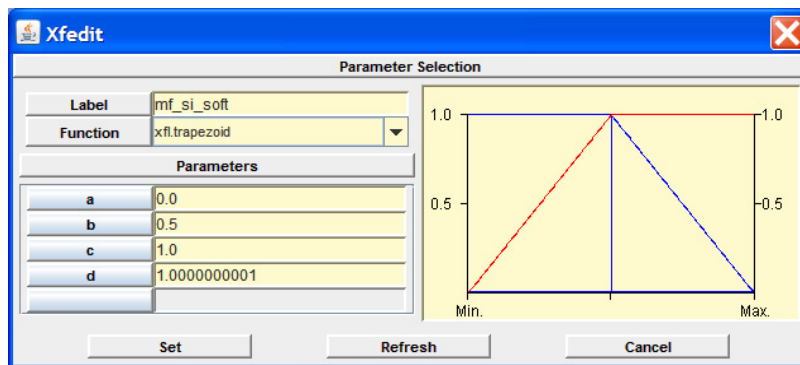


Figura 6-11. Ejemplo conjuntos borrosos *mf_si_soft* de la característica booleana *t_Arbol*.

6.1.2.1.1. Etiquetas lingüísticas

Como hemos visto, cada característica cualitativa booleana en la primera aproximación tiene como máximo dos etiquetas lingüísticas, una para C_j con valor *no* y otra con valor *sí*. Dependiendo de si se encuentran en las relaciones con los taxones, al menos una ocurrencia de los valores (3 y/o 2) para la etiqueta *sí* y los valores (-1 y/o 1) para la etiqueta *no*.

La segunda aproximación puede tener hasta un máximo de cuatro etiquetas lingüísticas *extremadamente_no*, *no*, *extremadamente_sí* y *sí*, su existencia depende de si se encuentra alguna ocurrencia de los valores -1, 1, 2 y 3 respectivamente, en las relaciones involucradas en dicha característica.

Las relaciones desconocidas no forman ningún conjunto borroso.

6.1.2.2. Características cualitativas aproximadas

Como vimos en el estudio del conocimiento:

En contraposición con las características cualitativas booleanas se detecta la existencia de características con respuestas aproximadas, se trata de cualidades que el usuario puede observar en mayor o menor grado, o incluso puede servirse de esta aproximación si no está seguro de estar observando plenamente dicha característica.

Estas características se aproximan gracias a las etiquetas lingüísticas genéricas poco, medio, mucho. Proporcionamos así una herramienta nueva al usuario, la posibilidad de aproximar la respuesta a una característica, podemos ver una imagen gráfica en la Figura 6-12.

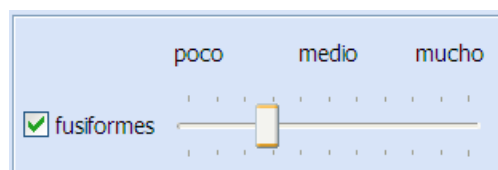


Figura 6-12. Posible implementación para características aproximadas.

Vamos a modelar conjuntos borrosos para representar este gradiente, los valores que toma el gradiente están comprendidos entre $[0, 1]$ y se comportan como el eje X de las funciones de pertenencia con valores comprendidos entre $[0,1]$. Para obtener el grado de pertenencia en cada punto, definimos cuatro conjuntos borrosos, los identificados por las etiquetas lingüísticas:

- Extremadamente poco

- Poco
- Medio
- Extremadamente mucho

Podemos ver la Figura 6-13, donde aparecen reflejadas las posibles funciones de pertenencia asociadas a estos conjuntos borrosos. El diseño de este modelo está fundamentado en los valores conocidos para las relaciones (taxón, característica) como veremos a continuación.

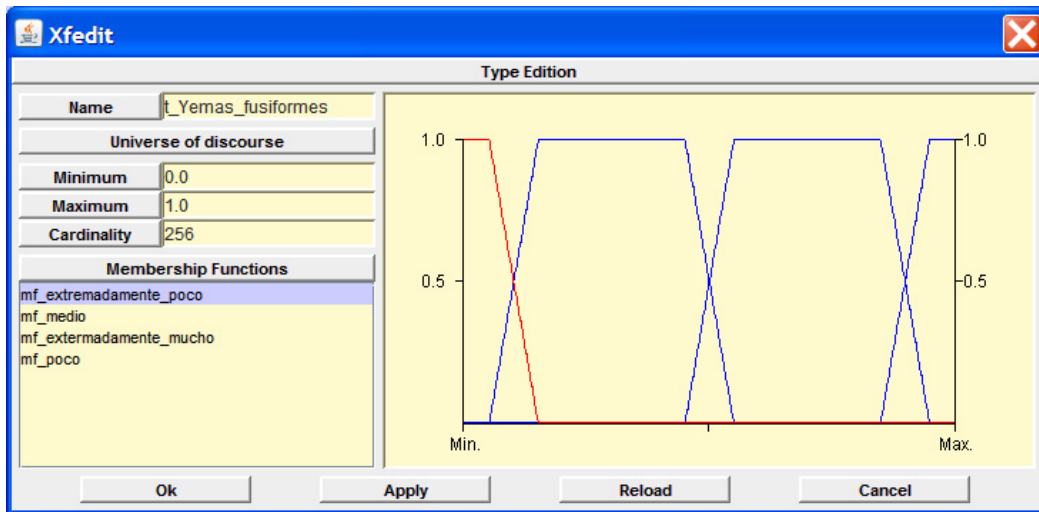


Figura 6-13. Ejemplo conjuntos borrosos en características aproximadas.

Los valores que conocemos para las relaciones de características cualitativas son (-1, 0, 1, 2, 3). Podemos modelar conjuntos borrosos para características aproximadas en función de estos valores. Partimos de una relación $R_{ij}(T_i, C_j)$ que tiene un valor V_{ij} comprendido en (-1, 0, 1, 2, 3). Los distintos significados de estos valores pueden ser observados en la siguiente Tabla 6-11 de nuevo.

Si vemos de nuevo la escala de la Figura 6-12 que proporciona los valores *poco*, *medio* y *mucho* para la característica *fusiformes*. Esta escala coincide con el eje X de las funciones de pertenencia diseñadas en la Figura 6-13, por lo que la escala toma valores numéricos entre [0, 1] que se corresponden automáticamente con un punto del eje X. Si observamos ahora las funciones de pertenencia, podemos ver el primer conjunto borrosos muy cercano al extremo izquierdo como el conjunto etiquetado por *extremadamente_poco*, este conjunto se identifica con los valores -1 (nunca se da) de las relaciones. De la misma manera los conjuntos *poco*, *medio* y *extremadamente_mucho* se corresponden con los valores 1 (poco frecuente), 2 (bastante frecuente) y 3 (siempre se

da) respectivamente. A continuación vemos el detalle de cada uno de ellos. Nótese que las relaciones evaluadas con 0 no intervienen, ya que no se dispone de información.

Tabla 6-11. Conjunto completo de posibilidades asociadas a una relación $R_{ij}(T_i, C_j)$.

-1	No es posible. Ese taxón quedaría rechazado si se da la característica
0	No hay información
1	En ocasiones se da. Es posible, aunque no muy frecuente
2	Frecuentemente, normalmente.
3	Siempre se da.

- El valor 1 indica que es poco frecuente que dicha característica C_j sea observada en un espécimen del taxón T_i . Si el usuario selecciona la característica C_j cerca del valor poco y tenemos el conjunto borroso formado por la etiqueta lingüística poco; podemos entonces identificar la correspondencia del conjunto con la relación R_{ij} . Es decir, para la característica C_j existe un conjunto borroso con función de pertenencia *mf_poco* que evalúa la premisa para el taxón T_i .

En el ejemplo de la Tabla 6-12, vemos la característica fusiforme, el taxón PINACEA tiene valor 1, por lo tanto es “poco frecuente” que se de dicha característica en este taxón. Si el usuario emite como respuesta un valor cercano a poco, la relación más probable es aquella marcada con 1, ya que las marcadas con 2 se dan muy frecuentemente y las marcadas con 3 se dan siempre. Por lo que si el usuario observa dicha característica solo un poco, nos estamos aproximando más a “Es posible, aunque no muy frecuente”, por lo tanto asociamos el conjunto borroso *mf_poco*.

De esta manera toda relación con valor 1, se evaluará con el conjunto borroso *mf_poco*.

Tabla 6-12. Ejemplo de característica aproximada, fusiformes.

Orden	“c” cuantitativa, “u” cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Grupo auto-excluyentes	Aproximadas	Orden relación proximadas	PINACEAE	Pseudotsuga
80		Yemas								
83			fusiformes			s	s		1	3

- El valor 2 indica que es bastante frecuente que dicha característica C_j sea observada en un espécimen del taxón T_i . Si el usuario selecciona la característica C_j cerca del valor medio y tenemos el conjunto borroso formado por la etiqueta

lingüística medio; podemos entonces identificar la correspondencia del conjunto con la relación R_{ij} . Es decir, para la característica C_j existe un conjunto borroso con función de pertenencia mf_medio que evalúa la premisa para el taxón T_i .

Si el taxón T_i tiene valor 2 indica que es “bastante frecuente” que se de dicha característica en este taxón. Si el usuario emite como respuesta un valor cercano a medio, la relación más probable es aquella marcada con 2, ya que las marcadas con 1 se dan poco frecuentemente y las marcadas con 3 se dan siempre. Por lo que si el usuario observa dicha característica medianamente, nos estamos aproximando más a “bastante frecuente”, por lo tanto asociamos el conjunto borroso mf_medio .

De esta manera toda relación con valor 2, se evaluará con el conjunto borroso mf_medio .

- El valor 3 indica que siempre se da dicha característica C_j en un espécimen del taxón T_i . Si el usuario selecciona la característica C_j cerca del valor mucho y tenemos el conjunto borroso formado por la etiqueta lingüística *extremadamente_mucho*; podemos entonces identificar la correspondencia del conjunto con la relación R_{ij} . Es decir, para la característica C_j existe un conjunto borroso con función de pertenencia $mf_extremadamente_mucho$ que evalúa la premisa para el taxón T_i .

En el ejemplo de la Tabla 6-12, vemos la característica fusiforme, el taxón *Pseudotsuga* tiene valor 3, por lo tanto dicha característica “siempre se da” en este taxón. Si el usuario emite como respuesta un valor cercano a mucho, la relación más probable es aquella marcada con 3, ya que las marcadas con 2 se dan muy frecuentemente y las marcadas con 1 se dan poco frecuentemente. Por lo que si el usuario observa dicha característica mucho, nos estamos aproximando más a “siempre se da”, por lo tanto asociamos el conjunto borroso $mf_extremadamente_mucho$.

De esta manera toda relación con valor 3, se evaluará con el conjunto borroso $mf_extremadamente_mucho$.

- El valor -1 indica que nunca se da dicha característica C_j en un espécimen del taxón T_i . Si el usuario selecciona la característica C_j en el extremo izquierdo del valor poco y tenemos el conjunto borroso formado por la etiqueta lingüística *extremadamente_poco*; podemos entonces identificar la correspondencia del conjunto con la relación R_{ij} . Es decir, para la característica C_j existe un conjunto

borroso con función de pertenencia *mf_extremadamente_poco* que evalúa la premisa para el taxón T_i .

Si el taxón T_i tiene valor -1 indica que “nunca se da” la característica en ese taxón. Si el usuario emite como respuesta un valor en el extremo izquierdo cerca de poco, nos estamos aproximando más a “nunca se da”, por lo tanto asociamos el conjunto borroso *mf_extremadamente_poco*.

De esta manera toda relación con valor -1, se evaluará con el conjunto borroso *mf_extremadamente_poco*.

Nótese la definición de 4 conjuntos borrosos para representación de las 4 etiquetas lingüísticas *extremadamente_poco*, *poco*, *medio* y *extremadamente_mucho*. Según la aproximación que indique el usuario, se obtiene un grado de pertenencia para los 4 conjuntos borrosos que representan dichas etiquetas. Cada especie tiene en la premisa de su regla una de estas etiquetas lingüísticas cuando así se haya encontrado en su relación (atendiendo a lo anteriormente descrito). Este es el típico ejemplo de etiqueta lingüística modelada con diferentes conjuntos borrosos descrito en el apartado 3.1.3, encontrado en numerosas aplicaciones de lógica, inferencia y control borroso.

6.1.2.2.1. Etiquetas lingüísticas

Como hemos visto, cada característica cualitativa aproximada, tendrá un número de etiquetas lingüísticas máximo igual a 4, *extremadamente_poco*, *poco*, *medio*, *extremadamente_mucho*, su existencia depende de si se encuentra alguna ocurrencia de los valores -1, 1, 2 y 3 respectivamente, en las relaciones involucradas en dicha característica.

Existen cuatro tipos de etiquetas lingüísticas, las que representan “Extremadamente poco de C_j ”, “Poco de C_j ”, “Medio de C_j ”, “Extremadamente mucho de C_j ”.

6.1.2.3. Características auto-excluyentes

Las características que encontramos en los grupos de auto-exclusión pueden ser de tres tipos:

- Booleanas
- Aproximadas
- Aproximadas relacionadas

Los dos primeros se comportan de forma idéntica a la descrita en los apartados inmediatamente anteriores. El tercer tipo sin embargo presenta peculiaridades que describiremos a continuación.

6.1.2.3.1. Características cualitativas auto-excluyentes relacionadas aproximadas

Como se describió en el apartado 4.2.3.3.1,

En los conjuntos de auto-excluyentes podemos encontrar características relacionadas las cuales tienen una “frontera” en la que ambas cualidades podrían darse y el observador podría no estar seguro de elegir una u otra. Si pensamos en poder dar al usuario un modo de aproximarse a estas características aunque no tenga certeza de cuál de ellas escoger como respuesta.

Una posible implementación de esta situación se observa en la Figura 6-14. El usuario puede eliminar su respuesta deseleccionando el botón check de la izquierda cumpliendo con los requisitos. Si desea evaluar dichas características no tiene más que acercar la barra al lado de la característica que observa.

Podemos representar estas dos características con dos conjuntos borrosos, de tal manera que las relaciones que tengan valores 2 y 3 para la característica verde oscuro, se identificarán por el conjunto borroso extremadamente muy verde oscuro, mientras que aquellas relaciones con valores -1 y 0 se representarán por el conjunto borroso extremadamente poco verde oscuro.

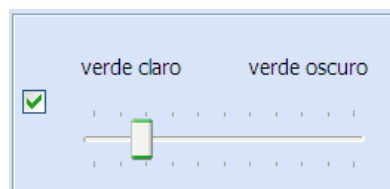


Figura 6-14. Posible implementación de características auto-excluyentes relacionadas aproximadas.

En la Tabla 6-13 tenemos un ejemplo de 3 características relacionadas aproximadas, el orden derivado de la información de la tabla sería:

Copa cónica o piramidal *Copa redondeada* *Copa aparasolada, densa*

En este caso tanto *Copa cónica o piramidal* como *Copa redondeada* son extremadamente poco de *Copa aparasolada, densa*.

Tabla 6-13. Ejemplo de grupo relacionado aproximado de 2 características.

Orden	"c" cuantitativa, " " cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Grupo auto-excluyentes	Aproximadas	Orden relación proximadas	Abies	Pinus halepensis	Pinus pinea
16			Forma de la Copa		1						
17			Copa cónica o piramidal			2	1	1	3	1	-1
18			Copa redondeada			2	1	2	-1	1	1
19			Copa aparasolada, densa.			2	1	3	-1	-1	2
20			Copa irregular			2			-1	1	-1

Si observamos la tabla anterior, es complicado fijar estos fuzzy sets, por ejemplo si tenemos tres conjuntos trapezoidales clásicos como los vistos en la Figura 6-15. Si el conjunto borroso más a la derecha se corresponde con la cualidad de "tener copa aparasolada, densa", todas las relaciones con la característica *Copa aparasolada, densa* que tengan valor 3 tendrán como premisa el grado de pertenencia a dicho conjunto borroso (más a la derecha). Aquellas relaciones con esta característica que tengan valores -1, tendrán como premisa el grado de pertenencia de la unión de los otros dos conjuntos borrosos. El valor 2 de la relación seguiría el mismo comportamiento que el 3, así como el valor 1 el de -1. Con esta interpretación perdemos información sobre las diferencias entre los valores (2, 3) y (-1, 1).

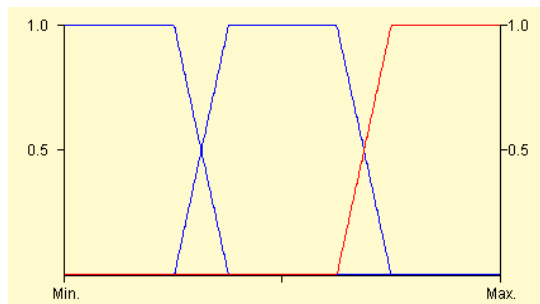


Figura 6-15. Conjuntos borrosos trapezoidales clásicos.

Observemos ahora el resultado de la unión de los dos conjuntos borrosos de más a la izquierda, utilizamos una t-conorma. Calculemos el siguiente ejemplo, si para el primer conjunto borroso tenemos valor 0.6 y para el segundo 0.4, entonces

- t-conorma Máximo $S(x,y) = \text{Max}(x,y) \rightarrow \text{Máximo}(0.4, 0.6) = 0.6$. Cabe destacar que esta es la menor de todas las t-conormas
- t-conorma Sum-Prod $S(x,y) = x + y - x \cdot y \rightarrow 0.4 + 0.6 - (0.4 \cdot 0.6) = 0.76$

- t-conorma dual de Lukasiewicz $W^*(x,y) = \min(1, x+y) \rightarrow \min(1, (0.4+0.6)) = 1$

Esto ocurre porque ambos conjuntos son complementarios, de tal manera en la t-conorma Lukasiewicz, se da la siguiente situación:

$$(A \cup \bar{A})(x) = \min(1, A(x) + (1 - A(x))) = 1 \quad (1)$$

Por lo tanto, para la unión de ambos conjuntos la t-conorma que mejor se comporta es la de Lukasiewicz, ya que el resto dan un grado menor que 1. Si el usuario ha seleccionado un valor muy alejado de *Copa aparasolada, densa*, y tenemos una relación con valor -1 para esa característica, todo el rango con grado 1 del conjunto borroso *extremadamente poco de Copa aparasolada, densa*, debería de ser continuo a medida que nos alejamos del conjunto borroso *extremadamente mucho de Copa aparasolada, densa*. Sin embargo vemos como para las t-conormas Máximo y Sum-Prod aparece un parón entre los dos subconjuntos borrosos cuya unión forma el conjunto borroso *extremadamente poco de Copa aparasolada, densa*.

Como hemos visto el mayor problema de esta solución es que al defuzzificar los valores perdemos la información adicional que nos proporcionan los valores 1 y 2. Además, para la unión de conjuntos en casos con más de dos características relacionadas, vemos que únicamente la t-conorma de Lukasiewicz se comporta correctamente.

6.1.2.3.1.1 Ampliación de la solución

Podemos plantear nuevos conjuntos borrosos más complejos que intenten cubrir estos problemas.

Si recordamos el punto anterior donde cada una de las características aproximadas se identificaban por cuatro conjuntos borrosos:

- Extremadamente poco
- Poco
- Medio
- Extremadamente mucho

Anteriormente para cada característica C_j obteníamos dos conjuntos borrosos, aquel en el que la característica se daba “extremadamente mucho de C_j ” y un segundo conjunto, “extremadamente poco de C_j ”, donde este último conjunto era la unión del resto de conjuntos “extremadamente mucho de C_k ”, para todas las características C_k pertenecientes al grupo de aproximación distintas de C_j . Podemos multiplicar el

número de conjuntos borrosos que hemos definido para conseguir, siguiendo el ejemplo anterior, tener cuatro conjuntos que se ajusten a los cuatro tipos de información de los que disponemos. Las características situadas a los extremos de la barra de selección, se acercarán al valor “extremadamente mucho de” esa característica, según se alejan hacia los extremos. Por el contrario, se acercarán al valor “extremadamente poco de” la característica, a medida que se acercan al centro. Las características centrales sin embargo, tienen en el centro de su área la zona “extremadamente mucho de” dicha característica, a medida que nos alejamos a ambos lados sobre el centro nos acercamos a la zona “extremadamente poco de”. Podemos ver un esquema de esta situación en la Figura 6-16.



Figura 6-16. Esquema de características relacionadas aproximadas, y sus áreas de mayor y menor grado de “se da” dichas características.

Cada una de las áreas delimitadas por una característica, puede dar lugar a tres conjuntos borrosos en vez del único del que antes disponíamos. Si antes nuestro conjunto era “Extremadamente mucho de”, ahora tenemos tres conjuntos más pequeños “Extremadamente mucho de”, “Medio de” y “Poco de”, los cuales podemos relacionarlos inmediatamente con los valores 3, 2 y 1 respectivamente para todas las relaciones entre un taxón y dicha característica. Si observamos la Figura 6-17, podemos ver como en el extremo izquierdo la característica *Copa cónica o piramidal*, está descrita por los tres conjuntos borrosos descritos en esa misma secuencia, es decir desde el extremo izquierdo hacia el centro, tal y como indicamos anteriormente. En la misma figura observamos en el extremo derecho el mismo caso para la característica *Copa aparasolada, densa*, sus conjuntos borrosos están secuenciados igualmente desde el extremo derecho hacia el centro.

Finalmente, en la zona central de la figura, observamos los conjuntos borrosos para la característica *Copa redondeada*, en este caso tenemos cinco conjuntos que debemos reducir a tres. Es sencillo observar que, existen dos conjuntos que identifican “Medio de”, situados a ambos lados del conjunto “Extremadamente de *Copa redondeada*”. A partir de estos dos conjuntos, obtenemos un conjunto unión de ambos, que será el conjunto borrosos “Medio de *Copa redondeada*”. El mismo proceso se llevará a cabo para el conjunto “Medio de *Copa redondeada*”.

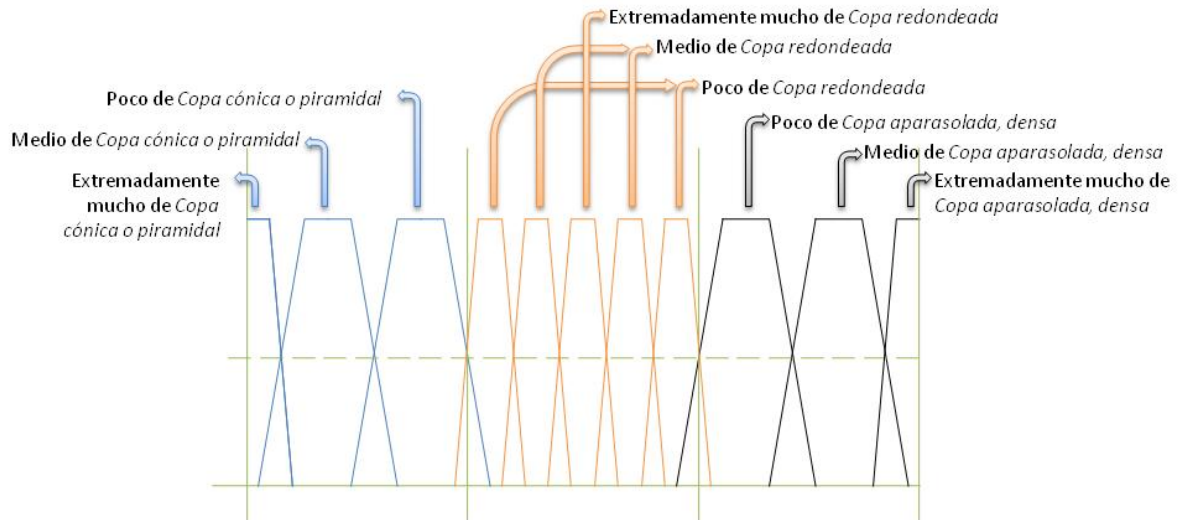


Figura 6-17. Conjuntos borrosos de tres características relacionadas aproximadas.

De esta manera, si tenemos una relación $R_{ij}(T_i, C_j)$ con la característica $C_j == \text{Copa redondeada}$, y dicha relación tiene valor 2. La regla del taxón T_i , tendrá como premisa el grado de pertenecía al conjunto borroso:

$$(\text{"Medio de } C_j") = ((\text{"Medio de } C_j \text{ izquierda}) \vee (\text{"Medio de } C_j \text{ derecha})) \quad (2)$$

El segundo problema que presenta la solución de un solo conjunto borroso, es que el conjunto borroso "Extremadamente poco de C_j ", es la unión del resto de conjuntos borrosos de las características del grupo de relaciones aproximadas excluyendo la propia C_j . La única t-conorma que da el resultado deseado es la de Lukasiewicz. Una posible solución sería incluir un conjunto borroso diseñado como la unión de los conjuntos mencionados, dando valor 1 a todo el rango deseado. Podemos ver en la Figura 6-18 dos ejemplos de esta solución.

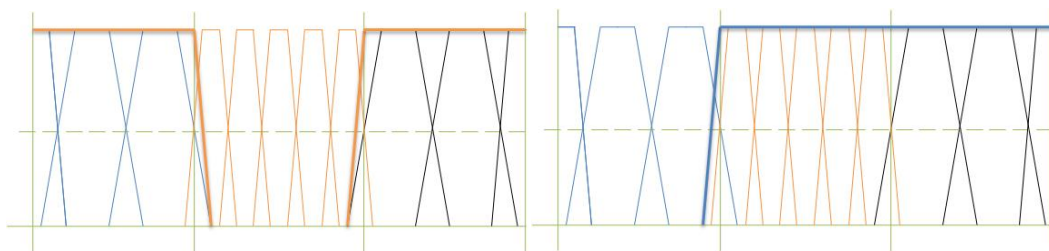


Figura 6-18. Conjuntos borrosos "Extremadamente poco de" para las características *Copa redondeada* y *Copa cónica o piramidal* en este orden.

6.1.2.3.2. Etiquetas lingüísticas

En las características auto-excluyentes se pueden dar características booleanas y aproximadas, las cuales generarán el mismo número de etiquetas lingüísticas que vimos con anterioridad.

Las características relacionadas aproximadas de la primera aproximación, generan dos etiquetas una para *Extremadamente mucho de* y otra para *Extremadamente poco de*. En este caso se tendrán en cuenta todas las etiquetas de todas las características, ya que los conjuntos de algunas características pueden estar generados mediante la unión de conjuntos de diferentes características componentes del mismo grupo de relación aproximada.

En la segunda aproximación, la situación es similar, excepto por que cada característica en vez de tener dos etiquetas tiene 4, *extremadamente_poco*, *poco*, *medio* y *extremadamente_mucho*.

6.2. Inferencia borrosa

6.2.1. Fuzzy set inferido

Cada uno de los taxones del conocimiento tendrá una regla de inferencia compuesta por tantas premisas como relaciones positivas y negativas tenga con cada una de las características relacionadas con el taxón. Las relaciones sin información no son premisas.

Todas las característica tienen asignadas una serie de variables lingüísticas modeladas mediante conjuntos borrosos con su función de pertenencia ajustada según su tipo de característica tal y como se ha estudiado.

Cada premisa de las reglas de inferencia de los taxones, se evalúa en función del grado de pertenencia al conjunto borroso asociado a la relación. Si tenemos una relación conocida $R_{ij}(T_i, C_j)$ con un valor V_{ij} donde C_j es una característica cualitativa, en función del valor de la relación elegiremos su correspondiente etiqueta lingüística de entre los conjuntos definidos para la característica C_j . En el caso en el que la relación conocida $R_{ij}(T_i, C_j)$ sea con una característica C_j cuantitativa, se evaluará el grado de pertenencia al conjunto borroso etiquetado para dicho taxón T_i , ya que los conjuntos definidos dependen de los límites informados para cada una de las relaciones independientemente.

Con el conjunto de premisas evaluadas con cierto grado de pertenencia a sus respectivos conjuntos borroso, se obtiene el grado de pertenencia al conjunto “soy del taxón T_i ”. Las reglas utilizan la conectiva conjunción AND -&- para el conjunto de todas las premisas. El cálculo de este grado de pertenencia se obtiene aplicando la norma triangular deseada. Se estudian las normas mínimo, producto y Lukasiewicz.

6.2.2. Fuzzy set fundamentado

Si únicamente utilizamos el grado inferido por las premisas evaluadas, para obtener el grado de pertenencia del espécimen de estudio a cada uno de los taxones. Podemos encontrarnos que la inferencia pueda situar en primer lugar un taxón como elección más probable, aunque el número de premisas evaluadas haya sido mínimo. Por ejemplo, tenemos un taxón con una única premisa evaluada cuyo grado es 1, lo que produciría que la regla infiriera automáticamente grado 1 de pertenencia a ese taxón, es decir máxima graduación. Paralelamente podríamos tener otro taxón con 10 premisas evaluadas y un grado inferido de 0,9. El conjunto solución resultante, situaría como taxón más probable al primero, pese a que la situación del segundo produce una determinación más fundamentada, porque se han evaluado 10 de sus premisas.

Para solventar este problema, se define un nuevo conjunto borroso llamado Conjunto Fundamentado, que determina la situación “muchas premisas evaluadas”. Para cada taxón se infiere el grado de fundamento en función del número de premisas evaluadas para ese taxón (llamémosle y) y el número total de características seleccionadas para el usuario (llamémosle x). El conjunto borroso viene definido por la función de pertenencia de la Figura 6-19, que obtiene un grado inferido igual a y/x .

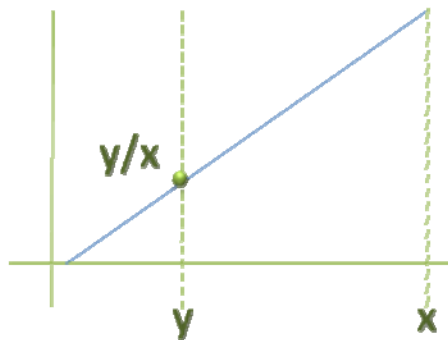


Figura 6-19. Función de pertenencia del conjunto borroso fundamentado.

6.2.3. Toma de decisiones

Sobre el universo de la Flora Ibérica $\mu: F \rightarrow [0, 1]$, hemos definido un conjunto borroso sobre cada taxón, Conjunto Preferencia, resultado de conectar a través de la conjunción AND, dos conjuntos borrosos:

Conjunto Preferencia = Conjunto Inferido \wedge Conjunto Fundamentado

Es decir, la conjunción entre el alto grado de pertenencia inferido con las características seleccionadas y el alto número de premisas evaluadas.

El cálculo del grado de pertenencia al Conjunto Preferido, se obtiene aplicando la t-triangular deseada. Se estudian las normas mínimo, producto y Lukasiewicz.

Finalmente obtenemos un conjunto de taxones posibles ordenado por mayor grado de pertenencia al Conjunto Preferencia. Esta solución es una conclusión imprecisa, si de este conjunto elegimos aquella con mayor grado de pertenencia estaremos facilitando una solución precisa, pero no podemos afirmar que sea la correcta. Es el experto usuario el que debe evaluar las posibles soluciones proporcionadas por el sistema, estudiar las premisas y el fundamento que han inferido la graduación, y establecer finalmente la solución más adecuada para la determinación taxonómica del espécimen en estudio.

Capítulo 7

Conclusiones

La determinación taxonómica de flora es un campo muy complejo de modelar y sistematizar por el alto volumen de información imprecisa que contiene. En este estudio se ha realizado un esfuerzo por normalizar la información contenida en las claves y descripciones recogidas en Flora Ibérica. Esta normalización ha derivado en un conjunto de variables con representaciones numéricas que forman parte de la entrada de conocimiento del sistema. La sistematización llevada a cabo consigue extraer información numérica del lenguaje científico.

Ante la dificultad de utilizar ordenaciones preestablecidas de las características a evaluar en una determinación. Se ha proporcionado al usuario herramientas a fin de que él mismo pueda realizar las observaciones en el orden deseado. Si bien la priorización de información a mostrar al usuario así como la restricción de datos inútiles o imposibles en los siguientes pasos de la determinación; es un área de estudio en la cual se puede ajustar el comportamiento aquí descrito.

El espectro del mundo de la flora es de una amplitud tal que se establece como requisito indispensable diseñar un sistema que sea auto-actualizable sencillamente por un experto; resultaría imposible abarcarlo dentro de este estudio. Con este fin, se ha diseñado una tabla de inserción de datos cómoda para un experto botánico, donde este mismo puede ampliar y modificar el conocimiento. El sistema es adaptativo, es capaz de utilizar esta información y actualizar de forma automática la visualización gráfica, la modelización de la información, la evaluación de las premisas y la inferencia del conjunto solución.

Si generalizamos todas las etapas que intervienen en el sistema, este modelo flexible, podría evolucionar a un 'determinador universal', que pudiera adaptarse a cualquier

otro tipo de sistemas de determinación distinto del aquí tratado, la botánica. Sería necesario generalizar la representación de la información y la inferencia de la solución, manejando tanto información nítida como borrosa.

Para realizar una primera evaluación de la normalización del conocimiento taxonómico, se ha implementado un sistema nítido a fin de utilizar la información ahí recogida y sentar las bases de un sistema borroso. Esta primera etapa nos ha permitido aumentar el nivel de información que podíamos extraer del conocimiento, gracias a lo cual hemos podido modelar un sistema más completo en todo el rango de información estudiada.

El sistema borroso reúne todos los requisitos obtenidos en el análisis del sistema nítido y aprovecha toda la información imprecisa que el sistema nítido no era capaz de procesar. Las características cualitativas con cualidades aproximadas son un claro ejemplo, hemos visto como el sistema borroso es capaz de proporcionar al usuario herramientas del lenguaje natural para emitir sus respuestas a estas características; en paralelo hemos relacionado estas respuestas con la información de probabilidad o frecuencia contenida en cada relación entre una característica y un taxón. Las características cuantitativas han suavizado también sus intervalos probables máximos, incluyendo cierto azar; no se elimina una posible solución si está cerca del intervalo dado.

Las etiquetas lingüísticas definidas han sido generalizadas de tal manera que, para cada característica cualitativa conocida se han establecido una serie de conjuntos borrosos genéricos. Los taxones implicados con dichas características son relacionados con el conjunto borroso más aproximado, atendiendo al valor de la relación entre el taxón y la característica. En ocasiones se ha detectado la necesidad utilizar la conectiva unión entre conjuntos borrosos para obtener el conjunto deseado. La definición de conjuntos borrosos genéricos produce una mayor adaptabilidad del sistema a futuras incorporaciones de conocimiento, así como para su extensión a otros campos de estudio.

El modo de evaluación de cada premisa se ha fijado obteniendo tres tipos de premisas. Excluyentes, las cuales eliminan automáticamente el taxón (en el sistema nítido de forma programática y en el sistema borroso por inferir grado de pertenencia 0). Indeterminadas, no se tiene información, no forman parte de las reglas. Positivas, su grado de pertenencia participa en la regla. Las reglas de inferencia de cada uno de los taxones se han modelado con estas premisas y operadores AND y OR, definiendo las t-normas y t-conormas a utilizar.

Por último, el resultado final no se da como solución única, si no como un conjunto de posibles soluciones ordenadas en función del grado de posibilidad inferido, como medio para la toma de decisión del experto. El usuario puede escoger de entre el conjunto de soluciones, aquel que como experto estime que es el más adecuado aunque no haya obtenido el grado máximo. Adicionalmente se han proporcionado medios para resolver la problemática de que el conjunto inferido pueda estar ordenado de forma poco “fundamentada”. Esto ocurre con taxones con pocas premisas evaluadas con alto grado de pertenencia, frente a otros taxones con un gran número de ellas. El mecanismo expuesto es un nuevo conjunto borroso en conjunción con el conjunto borroso inferido, que produce el conjunto borroso denominado preferencia. El conjunto fundamentado es aquel que evalúa la sentencia “muchas premisas seleccionadas”.

Capítulo 8

Trabajo Futuro

A lo largo de los distintos capítulos hemos detectado futuras evoluciones que se podrían realizar para mejorar la solución proporcionada. Adicionalmente, existen ámbitos muy interesantes que no han podido ser abordados en este estudio.

1. El sistema no ha podido ponerse a prueba en experimentos relevantes con expertos, por lo tanto no se han podido medir los resultados. El siguiente paso más inmediato sería realizar experimentos en los que expertos utilizaran el sistema.
 - a. Los experimentos más interesantes serían aquellos que nos permitieran comparar este sistema con otros sistemas en funcionamiento y métodos clásicos como el Dicotómico.
 - b. Podríamos realizar experimentos para medir y comparar con más exactitud los métodos propuestos entre sí, así como para ajustar diversos parámetros variables que hemos ido identificado. A raíz de los resultados obtenidos, podríamos ajustar estos parámetros variables, así como determinar la mejor solución en aquellos puntos en los que se han propuesto varias posibilidades. Sería necesario ampliar la implementación si se desea incluir más posibilidades.
2. La cantidad de características mostradas al usuario en cada punto, podría ser optimizada mediante un sistema experto, para evitar que el usuario pierda tiempo con características que no pueden aportar información decisoria para el conjunto solución en dicho punto. Por ejemplo, en cada momento, a partir de la lista de taxones posibles y sus relaciones con las características, si una

característica tiene el mismo valor en sus relaciones con todos los taxones que quedan como posibles, esa pregunta debería desaparecer ya que, aunque se conteste, no daría ninguna información. También podrían eliminarse aquellas características que no tengan relación conocida con los taxones actualmente posibles. Nótese, que aquellas que hayan sido contestadas no deben desaparecer, siempre existe la opción de eliminar o modificar la respuesta.

3. Aprendizaje automatizado a partir de ejemplo. Este aprendizaje actualizaría automáticamente elementos del sistema como el modelado de conjuntos borrosos, reglas, etc. Hemos visto que las definiciones de los conjuntos trapezoidales son muy genéricas, este sería un buen punto para aplicar aprendizaje con ejemplos.
4. Nuevo conocimiento. La inserción y modificación de conocimiento por parte del experto se realiza actualizando la información sobre la tabla Excel.
 - a. El sistema podría evolucionar a fin de incluir un asistente gráfico estructurado que guiara al usuario en la ampliación del conocimiento.
 - b. En el futuro se podría diseñar un procedimiento de aportación de nuevo conocimiento y validación del mismo por parte de un “consejo de expertos”. El repositorio de información se alimentaría con las nuevas aportaciones, se comprobaría su adecuación por el consejo y finalmente se podría poner a disposición de los usuarios por medio de actualizaciones periódicas.
5. Desarrollar una herramienta web para su uso online, que pueda ser utilizada por botánicos. El sistema podría almacenar en un repositorio, todos los datos incorporados en el conocimiento. También podría realizar estadísticas de acierto, los más buscados, las características con mayor número de evaluaciones etc. Estos datos podrían ser útiles a la hora de perfilar y evolucionar el sistema.
6. Diseño de un “determinador universal”. Objetivo ambicioso en el que se podrían generalizar los elementos del sistema para que este pudiera adaptarse a cualquier otro tipo de determinación, en otras disciplinas.
7. En este estudio hemos diseñado un sistema nítido y otro borroso. Se podrían diseñar nuevos sistemas utilizando otras técnicas del ámbito de la Inteligencia Artificial como por ejemplo redes neuronales, CBR, algoritmos genéticos, etc.

8. Evolucionar el sistema para incluir respuestas sí, no, no sé. En este estudio se han visto las complicaciones de esta posibilidad. La investigación de posibles soluciones a la problemática, se deja para futuros trabajos.
9. El conjunto de posibles taxones solución, podría ser acotado al inicio de la determinación por el propio usuario. El sistema podría proporcionar herramientas al usuario para partir de un taxón conocido. Por ejemplo, si sabe que el espécimen es de una familia determinada, buscaría la solución directamente dentro de esa familia.
10. El sistema podría mostrar a petición del usuario, para cada taxón componente del conjunto solución, la información de qué características se han visto involucradas en la inferencia del grado de pertenecía obtenido. Igualmente el sistema podría proporcionar información para los taxones descartados, acerca de que las características que han influido en su eliminación.
11. Se han estudiado características relacionadas aproximadas. Estas características cualitativas únicamente han sido detectadas en los grupos de auto-exclusión, pero es posible que se pueda dar fuera de estos grupos. A medida que aumente el conocimiento, se podrán detectar nuevas implicaciones que aporten información extra a la determinación.

Bibliografía

- [1] S. Castroviejo et al. (1997) *Flora Ibérica: Plantas Vasculares de la Península Ibérica e Islas Baleares*. Vol. V: Ebenaceae-Saxifragaceae. Real Jardín Botánico de Madrid. CSIC.
- [2] D. Dubois and H. Prade, *Fuzzy Sets and System: Theory and Applications*, Academic Press, New York, 1981.
- [3] J. C. Fodor, *On fuzzy implication operators*, *Fuzzy Sets and Systems* 41 (1991) 293--300.
- [4] E. Trillas and C. Alsina, *A reflection on what is a membership function*, *Mathw. Softcomput.* 6 (1999), pp. 201-215.
- [5] Xfuzzy. Pagina Web oficial. <http://www.imse.cnm.es>.
- [6] L.A. Zadeh, *Fuzzy sets*, *Inform. and Control* 8 (1965), pp. 338-353.
- [7] L.A. Zadeh, *The concept of linguistic variable and its application to approximate reasoning, parts I, II, III*, *Inform. Sci.* 8 (1975), pp. 199-249 8 (1975) 301-357; 9 (1975) 43-80
- [8] L.A. Zadeh *Fuzzy logic =Computing with words*. IEEE Trans. Fuzzy Syst., Vol. 4, (1996) pp. 103-111.
- [9] L.A. Zadeh, *From computing with numbers to computing with words – From manipulation of measurements to manipulation of perceptions*. In: P.P. Wang, Editor, *Computing with Words*, John Wiley & Sons (2001), pp. 35-68.
- [10] Flora Ibérica. Plantas vasculares de la Península Ibérica e Islas Baleares. <http://www.floraiberica.org/>.

- [11] Enlaces interesantes de botánica. <http://floramontiberica.org/enlaces.htm>
- [12] Sistema de información de Determinación Taxonómica de Flora Ibérica "DetFlora" recogido en www.fdi.ucm.es/profesor/lgarmend/
- [13] Asociación BIGA (Biodiversidad en Galicia). Plantae, espacio dedicado a información sobre flora vascular gallega. <http://www.biga.org/>.
- [14] T. Sánchez Velázquez. Clave taxonómica de la pteridoflora de las Islas Canarias. *Botanica Complutensis* 28: 39-50. 2004.
- [15] Herbario de la Universidad Pública de Navarra. <http://www.unavarra.es/servicio/herbario/htm/inicio.htm>
- [16] El Herbario Virtual del Mediterráneo Occidental. <http://herbarivirtual.uib.es/cas-med/index.html>
- [17] Proyecto Flora Ibérica "De lo real a lo imaginario", Universidad de Valencia. <http://www.florayfaunaiberica.com/>
- [18] Lecciones Hipertextuales de Botánica . <http://www.unex.es/polen/LHB/>
- [19] Los macrolíquenes epifitos de España. Una guía interactiva. <http://herbarivirtual.uib.es/cas-med/index.html>
- [20] Biology 211: Taxonomy of Flowering Plants. World Wide Flowering Plant Family Identification. <http://www.colby.edu/info.tech/BI211/PlantFamilyID.html>
- [21] Herbario virtual de las Islas Baleares. Prototipo de sistema de determinación. <http://130.206.78.170/fmi/iwp/cgi?-db=ClassBotanic&-loadframes>
- [22] A. Meskauskas and V. Gediminas. *An Expert system for integrating several manuals on determining plant diseases into one united database*. Technical University Lithuania. Instructional Technology Symposium 2001. http://www.ispp-itsymposium.org.nz/papers/submiss_1/
- [23] E. Gibaja Galindo. *Tesis Doctoral. Modelos de representación del conocimiento para la identificación taxonómica y aplicaciones*. Universidad de Granada 2004.
- [24] W. Fajardo Contrerasa, E. Gibaja Galindoa, A. Bailon Morillasb, P. Moral Lorenzoa. *An application of expert systems to botanical taxonomy*. *Expert Systems with Applications* 25 (2003) 425-430.

- [25] D. Díaz, L. Garmendia, A. Garmendia, A. Salvador. *Species Identification of specimens from Primula genus using fuzzy logic*. Segundo Congreso Internacional de Matemáticas en la Ingeniería y la Arquitectura. Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos, Universidad Politécnica de Madrid 2008, pages 175-182. ISBN: 978-84-7493-390-1.
- [26] D. Díaz, A. Garmendia, L. Garmendia, A. Salvador. *Diseño mediante lógica fuzzy de un sistema de determinación de especies de plantas de la familia Primula*. Journal of Mathematics & Design, volumen 9 nº 1, pages 7 – 22, 2008. ISSN 1515-7881
- [27] D. Díaz, A. Garmendia, L. Garmendia, A. Salvador. *Diseño mediante lógica fuzzy de un sistema de determinación de especies de plantas de la familia Primula*. 3ras. Jornadas de Matemáticas y Diseño M&D-2008. Publicación: En Actas. Argentina 2008.

Apéndice 1. Introducción a la lógica borrosa

La lógica clásica, la teoría de conjuntos clásica o la teoría de probabilidad pueden no ser adecuadas para tratar la imprecisión, la incertidumbre, la no especificidad, la vaguedad, la inconsistencia y la complejidad del mundo real. Esto motiva el nacimiento de los conjuntos difusos y las lógicas borrosas y explica su papel en la reestructuración de los fundamentos de las teorías científicas y sus aplicaciones, por lo que se están produciendo grandes avances tanto en áreas teóricas como en gran variedad de aplicaciones.

La teoría de la probabilidad sólo es capaz de representar uno de los tipos de incertidumbre que se basa en la aleatoriedad, no en la imprecisión de la información. *Lotfi A. Zadeh* en 1965 escribe su artículo en el que introduce una teoría sobre unos objetos, los conjuntos difusos, que son conjuntos de frontera no precisa y cuya función de pertenencia indica un grado. En la esfera de los predicados subjetivos, y por tanto imprecisos, la teoría de conjuntos clásica se enfrenta con obstáculos difíciles de superar.

Las lógicas borrosas necesitan generalizar las conectivas entre conjuntos borrosos. Los conectivos lógicos AND, OR y NOT, y las operaciones entre conjuntos intersección, unión y negación se generalizan respectivamente mediante normas triangulares, conormas triangulares y negaciones.

Cuando se pretende utilizar las relaciones borrosas para efectuar inferencias de razonamiento aproximado de la forma

Si x es P entonces y es Q

x es P'

y es Q'

se puede utilizar la regla composicional de inferencia de *Zadeh* y, en el caso de un universo en discurso, se obtendrán consecuencias lógicas en el sentido *Tarski*, pero se debe asegurar la propiedad de condicionalidad de la relación borrosa si queremos que se verifique el *Modus Ponens Generalizado* definido por *E. Trillas* [Trillas & Cubillo; 1996].

Parece necesario estudiar las propiedades algebraicas de las relaciones borrosas. Las más utilizadas son la reflexividad, simetría y T-transitividad. Una relación borrosa con estas tres propiedades es una **T-indistinguibilidad**, que generaliza a una relación de equivalencia. Sus aplicaciones son diversas, como la comparación y la clasificación, y es utilizada también en el aprendizaje inductivo automático. Es interesante estudiar las indistinguibilidades como complemento o negación de distancias, aprovechando los conocimientos sobre espacios métricos y espacios métricos generalizados.

Un tipo muy interesante de relaciones borrosas, casi siempre implicaciones lógicas, son los **T-preórdenes**, es decir, las relaciones reflexivas y T-transitivas. Su principal aplicación consiste en que un preorden define un operador de consecuencias en el sentido *Tarski* al aplicar la regla composicional de inferencia de *Zadeh*. Es fundamental el estudio y manejo de las relaciones residuadas de una norma triangular *T*, que al mismo tiempo son T-preórdenes y una cota superior de las relaciones T-condicionales.

TERNAS LÓGICAS

- **t-normas**

Según las aplicaciones se puede definir las operaciones de los conjuntos borrosos utilizando conectivos diferentes al mínimo, máximo y negación. Ya *B. Schweizer* y *A. Sklar* en "*Statistical Metric Spaces*", en 1960, trabajan las normas triangulares (o t-normas) mediante funciones generadoras, y observando sus propiedades, se comprueba que se pueden utilizar para generalizar la operación de intersección clásica así como las t-conormas para generalizar la unión clásica.

Las t-normas se comportan como conjunciones por lo que son ampliamente utilizadas en lógica borrosa. Las utilizamos por ello en la definición de expresiones de las

medidas de especificidad. Además las t-normas pueden ser utilizadas como generadoras de *modus ponens*. Los condicionales residuados asociados con cada una de las t-normas más importantes son bien conocidos. Por esto es preciso resumir en este apéndice sus propiedades más importantes.

Definición:

Una **norma triangular** (o brevemente una t-norma) es una operación binaria asociativa en $[0, 1]$, $T: [0, 1] \times [0, 1] \rightarrow [0, 1]$, que para todo x, y, z de $[0, 1]$ satisface:

T1) $T(x, 1) = x$, (elemento neutro 1)

T2) Si $x \geq x', y \geq y'$ entonces $T(x, y) \geq T(x', y')$ (monotonía)

T3) $T(x, y) = T(y, x)$ (simetría)

T4) $T(x, T(y, z)) = T(T(x, y), z)$ (asociatividad)

lo que significa que es asociativa, conmutativa, no decreciente y con una condición de contorno. Puede visualizarse como una superficie sobre el cuadrado unidad que contiene al segmento $(0,1,0)$ $(0,0,0)$, al segmento $(0,0,0)$ $(1,0,0)$ y al punto $(1,1,1)$. Esta definición puede generalizarse a conjuntos n -ários utilizando la propiedad asociativa, y a conjuntos infinitos. Entre las t-normas existe una relación de orden, y una relación de predominancia.

Las t-normas son muy utilizados en lógica borrosa para definir la intersección entre conjuntos borrosos, ya que generaliza la intersección clásica. Como operador lógico, son operadores que satisfacen la tabla lógica del conectivo lógico “y” (AND).

Una t-norma es **arquimediana** si y sólo si es continua y $T(x, x) < x$, para todo $x \in (0,1)$. Las t-normas producto y de Łukasiewicz son t-normas arquimedianas; la t-norma mínimo no lo es.

Una t-norma arquimediana es **estricta** si y sólo si es estrictamente creciente en $(0, 1) \times (0, 1)$. La t-norma producto es estricta. La t-norma de Łukasiewicz no lo es.

Una t-norma es **positiva** si para $x > 0$ e $y > 0$ se tiene que $T(x, y) > 0$. La t-norma mínimo y la t-norma producto son positivas. La t-norma de Łukasiewicz no lo es.

Se pueden definir las t-normas axiomáticamente o mediante funciones generadoras [B. Schweizer y A. Sklar. *Probabilistic Metric Space*. 1983].

- **t-conormas**

Las t-conormas, o conormas triangulares, son operadores $S: [0,1] \times [0,1] \rightarrow [0,1]$ muy utilizados en lógica borrosa para definir las uniones entre conjuntos borrosos pues generalizan la unión clásica. Como operador lógico, son operadores que satisfacen la tabla lógica del conectivo “o”, (OR).

Se pueden definir las t-conormas a partir de las t-normas, axiomáticamente o mediante funciones generadoras [B. Schweizer y A. Sklar. *Probabilistic Metric Space*. 1983].

Una operación $S: [0, 1] \times [0, 1] \rightarrow [0, 1]$ es una t-conorma si $T(x, y) = 1 - S(1-x, 1-y)$ es una t-norma.

Definición:

Dada una t-norma T , se define la conorma dual de T como

$$T^*(x, y) = 1 - T(1-x, 1-y).$$

La t-conorma dual del Mínimo es el Máximo, la dual del producto es la suma probabilística: $\text{Prod}^*(x, y) = x + y - xy$. La t-conorma dual de Łukasiewicz es $W^*(x, y) = \text{Mín}\{1, x+y\}$.

Definición:

Axiomáticamente se define una conorma triangular $S: [0, 1] \times [0, 1] \rightarrow [0,1]$, como un operador que satisface los cuatro axiomas siguientes.

S1) $S(x, 0) = S(0, x) = x$, para todo $x \in [0,1]$

S2) Si $x \geq x', y \geq y'$ entonces $S(x, y) \geq S(x', y')$ (monotonía)

S3) $S(x, y) = S(y, x)$ para todo $x, y \in [0,1]$ (simetría)

S4) $S(x, S(y, z)) = S(S(x, y), z)$ para todo $x, y, z \in [0,1]$ (asociatividad)

Una t-conorma es **arquimediana** si y sólo si es continua y $S(x, x) > x$, para todo $x \in (0, 1)$.

Una t-conorma arquimediana es **estricta** si y sólo si es estrictamente creciente en $(0, 1) \times (0, 1)$. El máximo no es arquimediana, Prod^* y W^* si lo son. Prod^* es una t-conorma arquimediana y estricta, y W^* no es estricta.

- **Negaciones**

Las **negaciones** son operadores $N: [0, 1] \rightarrow [0, 1]$ muy utilizados en lógica borrosa para definir tipos de complementos entre conjuntos borrosos que generalizan la negación clásica. Como operador lógico, son operadores que satisfacen la tabla lógica del “No”, (NOT).

Definición:

Una función $N: [0, 1] \rightarrow [0, 1]$ se define axiomáticamente como una negación si verifica los tres axiomas siguientes:

$$N1) N(0)=1$$

$$N2) N(1)=0$$

$$N3) N \text{ es no creciente.}$$

Una negación es **estricta** si y sólo si es continua y estrictamente decreciente. Una negación es **involutiva** si y sólo si $N(N(x)) = x$ para todo $x \in [0,1]$, es decir, si $N=N^{-1}$. Se dice que una negación es una **negación fuerte** si es continua, estrictamente decreciente e involutiva.

S. Weber demuestra el siguiente resultado: Sea T una t-norma (continua) y sea N una negación, entonces $S(x, y) = N^{-1}(T(N(x), N(y)))$ es una t-conorma (continua).

Sea T una t-norma arquimediana (estricta) generada por f , entonces $g = f \circ N$ genera una t-conorma arquimediana (estricta). Además $f(0) = g(1)$.

Definición: Conorma dual de T respecto de la negación N

Se define T^* como la t-conorma dual de T respecto de la negación N si:

$$T^*(x, y) = N(T(N(x), N(y))).$$

- **Familias de conectivos lógicos borrosos.**

Una **familia de conectivos lógicos borrosos** (T, S, N) está formada por una norma triangular T , una conorma triangular S y una negación N , y se denomina una **terna de De Morgan** o **terna de Morgan** cuando S es la t-conorma dual de T respecto a la negación N . Se utilizan para generalizar las operaciones de intersección, unión y complementario. En el caso clásico los conectivos lógicos dotan al conjunto de partes de un conjunto de una estructura de álgebra de *Boole*, pero esta estructura no se consigue en el caso borroso.

Hemos visto que la familia $\{\text{Mín}, \text{Máx}, 1-x\}$ propuesta por Zadeh no verifica el *tercio excluso* y la no contradicción ($\text{Mín}\{x, N(x)\} \neq 0$; $\text{Máx}\{x, N(x)\} \neq 1$), aunque verifica el resto de propiedades de un álgebra de Boole (como las leyes de Morgan).

Es sencillo comprobar que la distributividad implica la ley de absorción, que a su vez implica la idempotencia. Las normas y conormas arquimedianas no son idempotentes, y las sumas ordinales tampoco, por lo que $\{\text{Mín}, \text{Máx}, 1-x\}$ es la única familia continua que satisface la propiedad distributiva o modular. Por tanto la familia de Łukasiewicz con $N(x) = 1 - x$ no es distributiva aunque sí satisface la ley de no contradicción y el *tercio excluso*. Según las propiedades que interese que se verifiquen en las aplicaciones se elige una terna u otra.

Las familias de conectivos lógicos más utilizadas son:

	$T(x, y)$	Generador aditivo	$S(x, y)$	Generador aditivo	$N(x)$
Zadeh	$\text{Min}(x, y)$		$\text{Max}(x, y)$		$1-x$
	$x \cdot y$	$-\log(x)$	$x+y-xy$	$-\log(1-x)$	
$Yager_p$	$1-\text{Min}(((1-x)^p+(1-y)^p)^{1/p}, 1)$	$(1-x)^p$	$\text{Min}((x^p+y^p)^{1/p}, 1)$	x^p	$(1-x^p)^{1/p}$
$Dombi_\lambda$ $\lambda > 0$	$\frac{1}{1 + \left[\left(\frac{1}{x} - 1 \right)^{-\lambda} + \left(\frac{1}{y} - 1 \right)^{-\lambda} \right]}$	$\left(\frac{1-x}{x} \right)^\lambda$	$\frac{1}{1 + \left[\left(\frac{1}{x} - 1 \right)^\lambda + \left(\frac{1}{y} - 1 \right)^\lambda \right]^{1/\lambda}}$	$\left(\frac{x}{1-x} \right)^\lambda$	$1-x$
$Weber_\lambda$ $\lambda > -1$	$\text{Max}\left(\frac{x+y-1+\lambda xy}{1+\lambda}, 0 \right)$	$1 - \frac{\ln(1+\lambda x)}{\ln(1+\lambda)}$	$\text{Min}(x+y+\lambda xy, 1)$	$\frac{\ln(1+\lambda x)}{\ln(1+\lambda)}$	$\frac{1-x}{1+\lambda x}$
W Łukasiewicz: $Weber_\lambda$ con $\lambda=0$	$\text{Max}(x+y-1, 0)$		$\text{Mín}\{1, x+y\}$		$1-x$
$Hamacher_\gamma$ $\gamma > 0$	$\frac{xy}{\gamma + (1-\gamma)(x+y)}$	$\frac{1}{\gamma} \ln \frac{\gamma + (1-\gamma)x}{x}$	$\frac{x+y-xy-(1-\gamma)xy}{1-(1-\gamma)xy}$	$\frac{1}{\gamma} \ln \frac{\gamma + (1-\gamma)(1-x)}{1-x}$	$1-x$

RELACIONES BORROSAS

- **Estructura relacional borrosa**

Entre los predicados graduados pertenecientes a una misma variable lingüística usualmente existe una relación de antonimia. De hecho los valores lingüísticos de muchas variables pueden generarse a partir de un par de predicados antónimos y una serie de modificadores. Por ejemplo, para la variable lingüística *temperatura*, los predicados *frío* y *caliente* pueden generar los valores *helado*, *bastante frío*, *frío*, *bastante caliente*, *caliente*, *muy caliente* generados mediante los modificadores *muy* y *bastante*. El uso de conectivos lógicos de conjunción, disyunción y negación permiten combinar estas etiquetas para obtener otras como *templado* = *ni caliente ni frío*. La importancia de las relaciones de antonimia entre predicados se manifiesta en la cantidad de conocimiento que se adquiere gracias a la existencia de predicados antónimos. Cuando se transmite información mediante alguno de estos conceptos se supone de forma implícita la existencia del contrario y de una posición intermedia. Zadeh al definir variable lingüística utilizó gramáticas generativas de tal forma que, a partir de ciertos operadores, todos los términos lingüísticos de la variable están dados. De la misma forma si se tiene una semántica asociada a los términos lingüísticos básicos mediante conjunto borrosos se tiene una semántica asociada a todos los términos lingüísticos válidos, sin más que asociar a dichos operadores operaciones adecuadas entre conjuntos borrosos. Sin embargo consideramos que la asignación de un significado a esos términos mediante conjuntos borrosos se realiza de otra forma, con base en los términos implícitos existentes cuando se realiza una predicación, constituyendo una jerarquía de varios niveles, cada uno de ellos presentando diferente granularidad. La clase de conjuntos borrosos de cada nivel se obtendrá como una partición borrosa del universo compatible con una indistinguibilidad. Veamos como se definen estos conceptos.

Sea X un conjunto clásico. Una relación borrosa (o difusa) es una aplicación $R: X \times X \rightarrow [0, 1]$, es decir, una relación borrosa sobre X es un conjunto borroso sobre $X \times X$. Al conjunto (X, R) formado por un conjunto borroso X y una relación borrosa R se le llama **estructura relacional borrosa**. Entre las relaciones más destacadas en la teoría de conjuntos clásica tenemos las relaciones de orden y las de equivalencia, luego las propiedades que nos interesará definir ahora son la reflexiva, simétrica, antisimétrica y transitiva. Como con las operaciones entre conjuntos tenemos también muchas posibilidades y formas diferentes de definir las.

Se dice que una relación borrosa es **reflexiva** si $R(a, a) = 1$ para todo $a \in X$. Se dice que es **simétrica** si $R(a, b) = R(b, a)$ para todo $a, b \in X$. Es sencillo probar que una relación

borrosa es reflexiva si y sólo si las relaciones clásicas definidas por los subconjuntos de nivel de R son relaciones reflexivas. Lo mismo ocurre con la propiedad simétrica.

Una relación borrosa que es reflexiva y simétrica se denomina **relación de semejanza**. Un ejemplo de una relación de semejanza puede ser la representada por la matriz:

$$\begin{pmatrix} 1 & 0.3 & 0.7 \\ 0.3 & 1 & 0.4 \\ 0.7 & 0.4 & 1 \end{pmatrix}$$

Se dice que una relación borrosa es **α -reflexiva** (reflexiva al menos en un cierto grado), si $R(a, a)$ es siempre mayor o igual a un cierto valor α .

Debemos ser más cautelosos en el momento de definir la transitividad pues si a está cerca de b , y b está cerca de c , ¿podemos siempre asegurar que a está cerca de c ? Es de sobra conocido el comportamiento de los sinónimos respecto a la transitividad.

R transitiva \equiv Si $R(a, b)$ y $R(b, c)$ entonces $R(a, c)$

$$\equiv R(a, b) \wedge R(b, c) \leq R(a, c)$$

$$\equiv R(a, c) \geq \max\{\min\{R(a, x), R(x, c)\}\}$$

Y generalizando se dice que una relación borrosa es **T-transitiva** (T es una t-norma) si: $T(R(a, b), R(b, c)) \leq R(a, c)$ para todo $a, b, c \in X$.

- **Preórdenes e indistinguibilidades**

Una relación borrosa que sea reflexiva y T-transitiva se denomina un **T-preorden**. Una relación borrosa que sea reflexiva, simétrica y T-transitiva se dice que es una **T-similaridad** o una **T-indistinguibilidad**.

Los preórdenes, entre los que se encuentran las relaciones de implicación, son muy utilizados para realizar inferencias borrosas. La relación borrosa $J_A^T: X \times X \rightarrow [0, 1]$ definida por:

$$J_A^T(a, b) = \sup\{z: T(A(a), z) \leq A(b)\},$$

es un T-preorden.

Las indistinguibilidades generalizan a las relaciones de equivalencia clásicas y se utilizan para definir valores de “similitud” o distancias generalizadas. Se puede definir

partición borrosa y considerar la noción de compatibilidad de una partición respecto una T-indistinguibilidad.

LÓGICAS BORROSAS

La utilización usual del término “lógica borrosa” está ligado a una semántica de muy amplio espectro que es entendida básicamente como sinónimo de todo aquello que arranca del trabajo de *Zadeh* sobre conjuntos borrosos.

La lógica se ocupa de hacer inferencias verdaderas a partir de otras verdades. Como ciencia que estudia el razonamiento aproximado la lógica borrosa ha proporcionado un cálculo para las consecuencias imprecisas, resultado de gestionar convenientemente la vaguedad de las premisas y, en ocasiones, la fuerza de la implicación. Tanto la vaguedad de las premisas como la credibilidad de la conclusión se representan frecuentemente en términos de grados de verdad. La noción de grado de verdad puede interpretarse como una verdad parcial o como utilidad, esto es, una creencia subjetiva que tiene el agente en la verdad de esa proposición. Pero ahora nuestro grado de verdad no es sólo falso (0) o cierto (1) sino que puede haber grados. Se tiene por tanto que definir que se entiende por “premisa”, “consecuencia” y “conjetura”.

Uno de los problemas de la lógica borrosa es el tratamiento de la inferencia con información imprecisa y, en particular, el de la obtención de modelos para los enunciados condicionales, es decir, los del tipo: “*Si x es A entonces y es B* ”, donde x e y son elementos del universo y A y B son predicados vagos sobre el mismo. Con este fin se han definido operadores a partir de conceptos análogos de las lógicas bivaluada y multivaluada.

Un método de razonamiento borroso es un procedimiento de inferencia que deriva conclusiones entre un conjunto de reglas borrosas y un ejemplo. Para ello se tiene en cuenta el grado de compatibilidad, el grado de asociación, la función de ponderación, el grado de clasificación del ejemplo en cada una de las clases y la clasificación.

Es pues muy necesario proveer de teorías sobre las formas adecuadas de realizar estas inferencias, y de obtener consecuencias desde un conjunto de reglas dado.

- **Operador de consecuencias**

Sea E un conjunto cuyos elementos representan objetos lógicos, o proposiciones, y se parte de un conjunto de premisas V que será un subconjunto de E distinto del vacío. En este sentido *Alfred Tarski* axiomatizó la idea de “consecuencia” lógica a través de lo que se denomina **un operador de consecuencias** C , que es una aplicación de $\Phi(E)$ en $\Phi(E)$ que verifica:

- 1) $V \subset C(V)$ para todo $V \in \Phi(E)$ (las mismas premisas son consecuencias)
- 2) Si $V_1 \subset V_2$ entonces $C(V_1) \subset C(V_2)$ (a más premisas no menos consecuencias)
- 3) $C(V) = C(C(V))$ (se deducen todas las consecuencias posibles)

Luego, partiendo de un conjunto de premisas, mediante los operadores de consecuencias se obtiene el conjunto de consecuencias que puede deducirse por algún método adecuado. La primera propiedad nos dice que cada premisa puede considerarse una consecuencia, la segunda asegura la monotonía en el sentido de que al aumentar el número de premisas se aumenta el número de consecuencias, y la tercera dice que una vez halladas “todas” las consecuencias, ya no hay más consecuencias, pues $C^2 = C$. Una **lógica**, según la idea tarskiana, no es otra cosa que un conjunto de proposiciones provisto de un operador de consecuencias.

OPERADORES DE IMPLICACIÓN

A la hora de buscar generalizaciones las posibilidades son múltiples, pero debe mantenerse siempre que, al aplicar el nuevo concepto al caso límite de valores de pertenencia 0 o 1 el resultado coincida con el dado por el concepto original. Para representar enunciados condicionales los operadores de implicación deben buscarse entre aquellos que verifiquen los de la implicación lógica clásica. Por tanto los operadores de implicación generalizan la implicación lógica clásica, cuya tabla de verdad es

a	b	$a \Rightarrow b$
1	1	1
1	0	0
0	1	1
0	0	1

y una vez añadidas unas condiciones de monotonía propias del cálculo proposicional clásico. Puede definirse como:

$$a \Rightarrow b \equiv a' + b \equiv b + (a' \cdot b') \equiv a' + a \cdot b.$$

Siguiendo esta idea se puede definir la implicación borrosa como:

a	b	$a'+b$	$a \cdot b$	b	$a \cdot b + a' \cdot b'$	$a' \cdot b$	$a' \cdot b'$	a'
0	0	1	0	0	1	0	1	1
0	1	1	0	1	0	1	0	1
1	0	0	0	0	0	0	0	0
1	1	1	1	1	1	0	0	0

sin más que imponer que si la primera proposición es cierta y la segunda falsa entonces la implicación debe ser falsa. Pero también es interesante imponer que si ambas proposiciones son ciertas, la implicación debe ser cierta, con lo que nos quedamos únicamente con cuatro posibilidades: $a'+b$, $a \cdot b$, b , $a \cdot b + a' \cdot b'$. La primera posibilidad, a la que se llama **implicación material**, es la misma que se deduce de la tabla de la implicación lógica clásica. Recordemos que puede ser cualquier t-norma, + cualquier t-conorma y cualquier negación, según la terna de *Morgan* elegida. Cuando \cdot es el mínimo la implicación se denomina de *Mamdani* y es la más utilizada en control. También se usa mucho en control con buenos resultados la segunda: $a \cdot b$, sustituyendo además de por el mínimo, por otras t-normas. La tercera indica que al menos la implicación tenga el mismo valor de verdad que la segunda proposición. La cuarta es la equivalencia lógica, o doble implicación. Cada una de estas cuatro posibles implicaciones lógicas tienen propiedades diferentes y puede ser interesantes unas u otras según las aplicaciones.

En un Álgebra de Boole $(B, +, \cdot, ')$ se dice que una operación $\rightarrow: B \times B \rightarrow B$ es una implicación si, para todo x, y en B es $x \cdot (x \rightarrow y) \leq y$, desigualdad que es equivalente a $x \rightarrow y \leq x' + y$. Por tanto, la implicación material no es la única implicación, pero sí la mayor de todas. En la tabla anterior hemos visto que podemos tener seis funciones booleanas que cumplen dicha condición.

Los principales operadores de implicación son:

- **Implicación residuada**

Definición:

Se define la **implicación residuada** de una t-norma continua T , y se denota J^T , a la aplicación: $J^T: [0, 1] \times [0, 1] \rightarrow [0, 1]$ tal que:

$$JT(x, y) = \text{Sup } \{z: T(x, z) \leq y\}.$$

Dado un conjunto borroso sobre un universo entonces:

$$J^{\text{min}}(x, y) = \begin{cases} 1 & \text{si } x \leq y \\ y & \text{si } x > y \end{cases} \text{ es un Min-preorden borroso}$$

$$J^{\text{prod}}(x, y) = \text{Mín}(1, y/x) \text{ es un Prod-preorden borroso}$$

$$J^{\text{W}}(x, y) = \text{Mín}(1, 1-x+y) \text{ es un W-preorden borroso}$$

Se tienen pues tres familias de implicaciones lógicas borrosas que se deben estudiar. Cada una de ellas tiene sus ventajas e inconvenientes, por lo que en los sistemas expertos se utilizan una u otra según la conveniencia. Por ejemplo en un álgebra de Boole de probabilidades funciona bien la t-norma de Łukasiewicz.

- **S-Implicación**

Dada una t-conorma S y una negación N, se define una operación de S-implicación como $I(x, y) = S(N(x), y)$. Observamos que es una generalización de la implicación definida por $x' + y$.

Observamos que el concepto de implicación borroso no está unívocamente determinado. Además podemos generalizar también los conectivos lógicos y obtener nuevos tipo de implicación como:

- **QM-Implicación**

Dada una t-norma continua T, una S-conorma continua S y una negación N, se define una operación de QM-implicación como $I(x, y) = S(N(x), T(x, y))$ que es la forma de generalizar la expresión: $x' + x.y$.

- **Regla composicional de inferencia**

Dado una estructura relacional borrosa (E, R) y una t-norma T se denomina **transformada lógica** sobre (E, R) a $L_R^T: [0, 1] \rightarrow [0, 1]$ definida por:

$$LRT:[\mu](b) = \text{Sup}_{a \in E} T(\mu(a), R(a, b))$$

La transformada lógica es la aplicación de la regla composicional de inferencia de Zadeh y se utiliza para efectuar razonamiento aproximado de la forma

Si x es P entonces y es Q

x es P'

y es Q'

S. Cubillo demuestra en su memoria de doctorado los siguientes resultados para relaciones borrosas R reflexivas:

$$\mu \subseteq L_R^T: [\mu]$$

$$\mu_1 \subseteq \mu_2 \Rightarrow L(\mu_1) \subseteq L(\mu_2)$$

$$\mu \in T(E, R) \Leftrightarrow L(\mu) = \mu$$

Si R es un T -preorden y T es una t -norma continua entonces $L(L(\mu)) = L(\mu)$ y, por tanto, L_R^T es un operador de consecuencias.

Apéndice 2. Tabla Excel de conocimiento de determinación taxonómica.

Apéndice 2. Tabla Excel de conocimiento de determinación taxonómica.

Orden	"c" cuantitativa, " " cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Grupo auto-excluyentes	Aproximadas	Orden relación proximadas	PINACEAE	Abies	Abies pinsapo	Abies alba	Pseudotsuga	Pseudotsuga menziesii	Picea	Picea abies	Pinus	Pinus halepensis	Pinus uncinata	Pinus pinea	Pinus radiata	Pinus sylvestris	Pinus nigra	Pinus pinaster	Pinus nigra ssp. nigra	Pinus nigra ssp. salzmannii
2			TIPO DE TAXÓN		11	60			1	2	3	3	2	3	2	3	2	3	3	3	3	3	3	3	4	4
3			Número de familia						1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4			Número de género						0	1	1	1	2	2	3	3	4	4	4	4	4	4	4	4	4	4
5			Número de especie						0	0	2	1	0	3	0	4	0	11	15	10	9	14	13	12	13	13
6			Características de la planta a determinar, cualitativas y cuantitativas																							
7		Planta																								
8			Leñosa		2	15			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
9			Herbácea			15			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
10		Biotipo			2																					
11			Árbol (fanerófito)		1	1			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
12			Arbusto (nanofanerófito)			1			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
13			Matorral bajo (caméfito)			1			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
14	c		Altura (m)																							
15				Max						50	30	50	100	100	60	60	40	20	25	30	40	40	40	40	40	40
16		Forma de la Copa			1																					
17			Copa cónica o piramidal			2	1	1	1	3	3	3	0	0	0	0	1	1	1	-1	1	1	1	1	1	1
18			Copa redondeada			2	1	2	1	-1	-1	-1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
19			Copa aparasolada, densa.			2	1	3	1	-1	-1	-1	0	0	0	0	1	-1	-1	2	-1	-1	-1	-1	-1	-1
20			Copa irregular			2			1	-1	-1	-1	0	0	0	0	1	1	1	-1	1	1	1	1	1	1
21		Tallo																								
22			Tallo articulado			3	2	2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
23			Tallo no articulado			3	2	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
24			Resinoso						3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
25			No resinoso						-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
26		Tronco			1																					
27			Tronco profundamente hendido						1	0	0	0	0	0	0	0	1	0	0	0	0	3	0	0	0	0
28			Tronco sólo ramificado en la parte superior						1	0	0	0	0	0	0	0	1	0	0	3	0	0	0	0	0	0
29			Tronco de color pardo rojizo						1	0	0	0	0	0	0	0	1	0	0	0	0	3	0	0	0	0
30			Tronco tortuoso			4			1	0	0	0	0	0	0	0	1	2	0	0	-1	0	0	0	0	0
31			Tronco Recto			4			1	0	0	0	0	0	0	0	1	1	0	0	3	0	0	0	0	0

Apéndice 2. Tabla Excel de conocimiento de determinación taxonómica.

Orden	"c" cuantitativa, " u" cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Grupo auto-excluyentes	Aproximadas	Orden relación proximadas	PINACEAE	Abies	Abies pinsapo	Abies alba	Pseudotsuga	Pseudotsuga menziezii	Picea	Picea abies	Pinus	Pinus halepensis	Pinus uncinata	Pinus pinea	Pinus radiata	Pinus sylvestris	Pinus nigra	Pinus pinaster	Pinus nigra ssp. nigra	Pinus nigra ssp. salzmannii
32			Ritidoma (parte exterior de la corteza, que se desprende)		1																					
33			Delgado y liso			5			1	0	0	0	0	0	0	0	1	3	3	-1	-1	3	0	-1	0	0
34			grueso			5			1	0	0	0	0	0	0	0	1	-1	-1	3	3	-1	0	3	0	0
35			ceniciento-plateado en los ejemplares jóvenes						1	0	0	0	0	0	0	0	1	3	0	-1	-1	-1	3	0	3	3
36			castaño oscuro en los adultos			6			1	0	0	0	0	0	0	0	1	-1	-1	3	0	-1	3	0	3	3
37			negruzco, grisáceo o pardo grisáceo			6			1	0	0	0	0	0	0	0	1	0	3	-1	1	-1	1	3	1	1
38			anaranjado o pardo rojizo			6			1	0	0	0	0	0	0	0	1	-1	-1	-1	3	3	-1	-1	-1	-1
39			Ramificación		1																					
40			Ramificación monopódica			7			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
41			Ramificación simpódica			7			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
42			Ramificación verticilada regular			7			1	3	3	3	-1	-1	3	3	2	2	2	2	2	2	2	2	2	2
43			Ramificación verticilada irregular			7			1	-1	-1	-1	3	3	-1	-1	1	1	1	1	1	1	1	1	1	1
44			Guía y puntas de las ramas recurvadas, colgantes			8			1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
45			Guía erecta y puntas de las ramas no recurvadas			8			1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
46			Ramillas (brotes del año)		2																					
47			ramillas redondeadas			9			1	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48			ramillas cuadrangulares			9			0	-1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
49			ramillas comprimidas			9			0	-1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
50			ramillas dispuestas en un plano			10			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
51			ramillas no dispuestas en un plano			10			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
52			ramillas delgadas			11			1	0	0	0	0	0	0	0	1	3	-1	3	0	3	0	0	0	0
53			ramillas gruesas			11			1	0	0	0	0	0	0	0	1	-1	3	-1	0	-1	0	0	0	0
54			ramillas con apófisis decurrentes, prominentes y separadas por surcos			12			1	-1	-1	-1	0	0	3	3	0	0	0	0	0	0	0	0	0	0
55			ramillas con apófisis poco prominentes			12			1	0	0	0	3	3	-1	-1	0	0	0	0	0	0	0	0	0	0
56			ramillas densamente pubescentes, el primer año			13			1	1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
57			ramillas glabras, nada pruinosas.			13			1	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58			ramillas poco pruinosas			13			1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
59			ramillas cenicientas			14			1	1	0	3	0	0	0	0	1	3	-1	-1	-1	-1	-1	-1	-1	-1
60			ramillas rojizas, amarillas o acastañadas			14			1	1	3	-1	0	0	0	0	1	-1	3	3	3	3	3	3	3	3

Apéndice 2. Tabla Excel de conocimiento de determinación taxonómica.

Orden	"c" cuantitativa, "u" cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Grupo auto-excluyentes	Aproximadas	Orden relación proximadas	PINACEAE	Abies	Abies pinsapo	Abies alba	Pseudotsuga	Pseudotsuga menziesii	Picea	Picea abies	Pinus	Pinus halepensis	Pinus uncinata	Pinus pinea	Pinus radiata	Pinus sylvestris	Pinus nigra	Pinus pinaster	Pinus nigra ssp. nigra	Pinus nigra ssp. salzmannii
61			Disposición de las hojas																							
62			Presencia de braquiblastos (tallos de crecimiento limitado, con hojas)	3	59				1	-1	-1	-1	-1	-1	-1	-1	3	3	3	3	3	3	3	3	3	
63			Sin braquiblastos		59				1	3	3	3	3	3	3	3	-1	-1	-1	-1	-1	-1	-1	-1	-1	
64			Hojas alternas		16				3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
65			Hojas opuestas o verticiladas (de otra manera)		16				-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
66			Hojas decusadas (opuestas, en cruz con anteriores y siguientes)		16				-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
67			Hojas en verticilos de 4		16				-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
68			Hojas imbricadas																							
69			Braquiblastos	3																						
70			Todas las hojas en braquiblastos (tallos de crecimiento limitado)		16				1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	
71			Hojas en macroblastos (tallo normal) y en braquiblastos																							
72			Braquiblastos con hojas fasciculadas						1	-1	-1	-1	-1	-1	-1	-1	1	3	3	3	-1	3	3	3	3	
73			Braquiblastos rudimentarios, muy cortos, con 2 hojas		17				1	-1	-1	-1	-1	-1	-1	-1	1	3	2	3	1	3	3	3	3	
74			Braquiblastos rudimentarios, muy cortos, con 3 hojas		17				1	-1	-1	-1	-1	-1	-1	-1	1	-1	1	-1	2	-1	-1	-1	-1	
75			Braquiblastos bien desarrollados con muchas hojas		17				1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
76			macroblastos con hojas escuamiformes sin clorofila		16				1	-1	-1	-1	-1	-1	-1	-1	3	3	3	3	3	3	3	3	3	
77			macroblastos con hojas en espiral		16				2	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
80			Yemas																							
81			agudas		18	3	1		1	1	-1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	
82			obtusas		18	3	2		1	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
83			fusiformes			s	s		1	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	
84			resinosas		19	4	2		1	1	3	-1	0	0	0	0	1	0	3	0	0	3	3	-1	3	
85			no resinosas		19	4	1		1	1	-1	3	0	0	0	0	1	0	-1	0	0	-1	-1	3	-1	
86			con escamas recurvadas en la punta		20				1	0	0	0	0	0	0	0	1	0	-1	0	0	-1	-1	3	-1	
87			con escamas aplicadas		20				1	0	0	0	0	0	0	0	1	0	3	0	0	3	3	-1	3	
88			Presencia de hojas																							
89			Hojas verdes, asimiladoras	4	21				3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
90			Hojas pardas, no asimiladoras (sin hojas verdes)		21				-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	

Apéndice 2. Tabla Excel de conocimiento de determinación taxonómica.

Orden	"c" cuantitativa, " cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Grupo auto-excluyentes	Aproximadas	Orden relación proximadas	PINACEAE	Abies	Abies pinsapo	Abies alba	Pseudotsuga	Pseudotsuga menziesii	Picea	Picea abies	Pinus	Pinus halepensis	Pinus uncinata	Pinus pinea	Pinus radiata	Pinus sylvestris	Pinus nigra	Pinus pinaster	Pinus nigra ssp. nigra	Pinus nigra ssp. salzmannii
91			Color de hojas		4																					
92			Color verde oscuro			22			1	1	-1	0	0	0	3	3	1	-1	3	3	0	-1	-1	3	-1	-1
93			Color verde claro			22			1	0	0	0	0	0	-1	-1	1	3	-1	-1	0	-1	3	-1	3	3
94			Color más o menos glauco			22			1	1	3	0	0	0	-1	-1	1	0	-1	0	0	3	-1	-1	-1	-1
95			Forma de las hojas		4																					
96			Hojas escumiformes			23			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
97			Hojas aciculares			23			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
98			hojas lineares aplanadas			23			1	3	3	3	0	0	3	3	1	1	1	1	1	1	1	1	1	1
99			hojas de sección tetragonal o subtetrales			23			1	1	1	1	0	0	3	3	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
100			agudas y/o mucronadas						1	0	0	0	0	0	0	0	1	0	3	0	0	3	0	3	0	0
101			ligeramente emarginadas			25			1	1	-1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
102			no emarginadas			25			1	1	3	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
103	c		Longitud de las hojas (mm)																							
104				Max						30	16	30	35	35			250	150	80	200	##	70	160	250	##	##
105				Min						6	6	10	20	20			30	60	40	100	70	30	60	100	80	60
106				MaxInt														120		150						
107	c		Anchura de las hojas (mm)																							
108				Max						3	2,5	2						1	2	1,5	2	2	2	2	2	2
109				Min						2	1,5	2						0,7	1	1	1	1	0,8	1	2	1
110			Características de hojas		4																					
111			Hojas caducas			24			1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
112			Hojas perennes			24			2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
113			Hojas flexibles			26	1	1	1	1	-1	3	3	3	0	0	1	3	-1	2	3	-1	1	-1	-1	3
114			Hojas rígidas			26	2	2	1	1	3	-1	-1	-1	0	0	1	-1	3	1	-1	3	1	3	3	-1
115			Hojas discoloras			27			1	1	-1	3	0	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
116			Hojas concoloras			27			1	1	3	-1	0	0	0	0	3	3	3	3	3	3	3	3	3	3
117			con dos bandas estomáticas en la cara inferior						1	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
118			Con una capa de células hipodérmicas			28			1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	-1	3
119			con dos a cinco capas de células hipodérmicas			28			1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3	-1
120			cicatriz de la hoja circular						1	3	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0
121			con canales resiníferos			5			1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

Apéndice 2. Tabla Excel de conocimiento de determinación taxonómica.

Orden	"c" cuantitativa, " " cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Grupo auto-excluyentes	Aproximadas	Orden relación proximadas	PINACEAE	Abies	Abies pinsapo	Abies alba	Pseudotsuga	Pseudotsuga menziesii	Picea	Picea abies	Pinus	Pinus halepensis	Pinus uncinata	Pinus pinea	Pinus radiata	Pinus sylvestris	Pinus nigra	Pinus pinaster	Pinus nigra ssp. nigra	Pinus nigra ssp. salzmannii
122			Canales resiníferos		5																					
123			canales resiníferos marginales o submarginales			29			1	1	1	1	3	3	0	0	0	3	0	0	0	3	-1	-1	-1	-1
124			canales resiníferos centrales			29			1	1	1	1	-1	-1	0	0	1	-1	0	0	0	-1	3	3	3	3
125	c		Número de canales resiníferos																							
126				Max						2	2	2						8					17	18	17	17
127				Min						2	2	2						3					3	2	3	3
128			Sexo																							
129			Monoico (plantas con flores macho y flores hembra)			30			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
130			Dioico (plantas macho y plantas hembra)			30			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
131			Hermafrodita (flores macho y hembra a la vez)			30			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
132			Flores																							
133			Flores con estambres, pistilo o ambas cosas						-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
134			Flores en inflorescencias con escamas (conos)		6	31			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
135			Flores no agrupadas en conos			31			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
136			Frutos																							
137			Frutos carnosos			32			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
138			Frutos leñosos (piñas o estróbilos)		11	32			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
139			Conos masculinos		6																					
140			axilares			33			1	3	3	3	0	0	1	1	0	0	0	0	0	0	0	0	0	0
141			terminales			33			1	-1	-1	-1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
142			agrupados en la base de los brotes anuales			34			1	-1	-1	-1	-1	-1	-1	-1	3	3	3	3	3	3	3	3	3	3
143			solitarios en el centro de los fascículos de hojas			34			1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
144			numerosos en la cara inferior de las ramas						1	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
145			dos sacos polínicos por escama			35			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
146			3-7 sacos polínicos por escama			35			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Apéndice 2. Tabla Excel de conocimiento de determinación taxonómica.

Orden	"c" cuantitativa, " u cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Grupo auto-excluyentes	Aproximadas	Orden relación proximadas	PINACEAE	Abies	Abies pinsapo	Abies alba	Pseudotsuga	Pseudotsuga menziezii	Picea	Picea abies	Pinus	Pinus halepensis	Pinus uncinata	Pinus pinea	Pinus radiata	Pinus sylvestris	Pinus nigra	Pinus pinaster	Pinus nigra ssp. nigra	Pinus nigra ssp. salzmannii
147			Conos femeninos (estróbilos)		11																					
148			Estróbilos erectos			36			1	3	3	3	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
149			Estróbilos colgantes, patentes o reflejos			36			1	-1	-1	-1	3	3	3	3	3	3	3	3	3	3	3	3	3	
150			Estróbilos sólo en ramas superiores de la copa			37			1	3	3	3	-1	-1	-1	-1	0	0	0	0	0	0	0	0	0	
151			Estróbilos en toda la copa			37			1	-1	-1	-1	3	3	3	3	0	0	0	0	0	0	0	0	0	
152			Estróbilos en verticilos de 2 a 5.						1	0	0	0	0	0	0	0	1	1	-1	-1	3	-1	-1	-1	-1	
153			cilíndricos			38			1	3	3	3	0	0	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	
154			doliiforme (forma de barril)			38			1	-1	-1	-1	0	0	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	
155			globoso elipsoidal u ovoide (base más ancha que el ápice)			38			1	-1	-1	-1	0	0	0	0	1	1	1	3	3	1	1	1	1	
156			Oblongo-cónicos (alargados)			38			1	-1	-1	-1	0	0	0	0	1	3	-1	-1	-1	-1	-1	3	-1	
157			Ovoideo (forma de huevo) - cónicos			38			1	-1	-1	-1	0	0	0	0	1	-1	3	-1	1	3	3	1	3	
158			Estróbilos de maduración anual			39			1	0	0	0	0	0	3	3	-1	-1	-1	-1	-1	-1	-1	-1	-1	
159			Estróbilos de maduración bi o trienal			39			1	0	0	0	0	0	-1	-1	3	3	3	3	3	3	3	3	3	
160			Estróbilos muy asimétricos						1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	-1	3	-1	-1	-1	-1	
161			Estróbilos con pedúnculo evidente, persistentes muchos años			40			1	0	0	0	0	0	0	0	1	3	-1	-1	0	-1	-1	-1	-1	
162			Estróbilos sésiles o subsésiles, más o menos caducos			40			1	0	0	0	0	0	0	0	1	-1	3	3	0	3	3	3	3	
163			Estróbilos brillantes			41			1	0	0	0	0	0	0	0	1	0	3	3	0	-1	0	3	0	
164			Estróbilos mate			41			1	0	0	0	0	0	0	0	1	0	-1	-1	0	3	0	-1	0	
165			color castaño o pardo amarillento en la madurez						1	0	0	0	0	0	0	0	1	3	3	3	0	3	3	3	3	
166			Estróbilos indehiscentes durante varios años			42			1	0	0	0	0	0	0	0	1	1	-1	-1	0	-1	-1	2	-1	
167			Estróbilos marcescentes (abiertos) durante varios años.			42			1	0	0	0	0	0	0	0	1	3	-1	-1	0	-1	-1	1	-1	
168			Estróbilos caducos			42			1	0	0	0	0	0	0	0	1	-1	3	3	0	3	3	-1	3	
169	c		Longitud de los estróbilos (cm)																							
170				Max					22	20	17	20	10	10	18	18	22	12	7	15	14	6	8	22	8	
171				Min						10	10	10	5	5	10	10	3	6	5	8	7	3	4	8	5	
172	c		Anchura de los estróbilos (cm)																							
173				Max						4	4	4	3,5	3,5			10	4,5	3	10	8	4	4	8	4	
174				Min						3	3	3	2,5	2,5			2	3,5	2	7	5	2	2	5	2	

Apéndice 2. Tabla Excel de conocimiento de determinación taxonómica.

Orden	"c" cuantitativa, " " cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Grupo auto-excluyentes	Aproximadas	Orden relación proximadas	PINACEAE	Abies	Abies pinsapo	Abies alba	Pseudotsuga	Pseudotsuga menziesii	Picea	Picea abies	Pinus	Pinus halepensis	Pinus uncinata	Pinus pinea	Pinus radiata	Pinus sylvestris	Pinus nigra	Pinus pinaster	Pinus nigra ssp. nigra	Pinus nigra ssp. salzmannii
175			Escamas de los estróbilos		11																					
176			Estróbilos con cuatro escamas a modo de valva.			43			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
177			Estróbilos con 6 o más escamas			43			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
178			Escamas opuestas o verticiladas			44			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
179			Escamas en espiral			44			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
180			Escamas caducas al madurar			45			1	3	3	3	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
181			Escamas persistentes			45			1	-1	-1	-1	3	3	3	3	3	3	3	3	3	3	3	3	3	
182			escamas libres			46			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
183			escamas concrecentes (soldadas)			46			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
184			Escamas planas a modo de valvas			60			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
185			Escamas peltadas			60			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
186			con el ápice recurvado			47			1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
187			con el ápice no recurvado			47			1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
188			presencia de escamas esteriles		8				2	2	2	2	3	3	3	3	2	2	2	2	2	2	2	2	2	
189			escamas fértiles en axila de esteriles						3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
190			con la parte externa (apófisis) con una protuberancia (ombiligo)		7				1	-1	-1	-1	-1	-1	-1	-1	3	3	3	3	3	3	3	3	3	
191	c		Número de semillas por escama																							
192				Max					2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
193				Min					2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
194			Apófisis		7																					
195			Apófisis de las escamas basales más prominentes que el resto			48			1	0	0	0	0	0	0	0	1	-1	-1	-1	3	-1	-1	-1	-1	
196			Apófisis de todas las escamas más o menos similares			48			1	0	0	0	0	0	0	0	1	3	3	3	-1	3	3	3	3	
197			Apófisis muy prominentes, ganchudas y retrorsas.						1	0	0	0	0	0	0	0	1	-1	3	-1	-1	-1	-1	-1	-1	
198			con ombiligo punzante (prominente y más o menos piramidal)			49			1	0	0	0	0	0	0	0	1	-1	-1	-1	3	-1	-1	3	-1	
199			con ombiligo no punzante (plano o poco prominente)			49			1	0	0	0	0	0	0	0	1	3	3	3	-1	3	3	-1	3	

Apéndice 2. Tabla Excel de conocimiento de determinación taxonómica.

Orden	"c" cuantitativa, " u cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Grupo auto-excluyentes	Aproximadas	Orden relación proximadas	PINACEAE	Abies	Abies pinsapo	Abies alba	Pseudotsuga	Pseudotsuga menziesii	Picea	Picea abies	Pinus	Pinus halepensis	Pinus uncinata	Pinus pinea	Pinus radiata	Pinus sylvestris	Pinus nigra	Pinus pinaster	Pinus nigra ssp. nigra	Pinus nigra ssp. salzmannii
200			Escamas estériles (tectrices) de los estróbilos		8																					
201			exertas			50			1	1	-1	3	3	3	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
202			inclusas			50			1	1	3	-1	-1	-1	3	3	3	3	3	3	3	3	3	3	3	
203			trilobadas en el ápice			51			1	0	0	0	3	3	-1	-1	0	0	0	0	0	0	0	0	0	
204			no lobadas			51			1	0	0	0	-1	-1	3	3	0	0	0	0	0	0	0	0	0	
205			con el limbo cuspidado			52			1	1	-1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	
206			con el limbo mucronado			52			1	1	3	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	
207			diminutas						1	-1	-1	-1	-1	-1	-1	-1	3	3	3	3	3	3	3	3	3	
208			Semillas																							
209			Plantas sin semillas ni flores (con esporas)			53			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
210			Plantas con semillas		10	53			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
211			Rudimentos seminales al descubierto			54			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
212			Rudimentos seminales encerrados en un ovario			54			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
213			Características de las semillas		10																					
214			Semillas solitarias						-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
215			semillas en cara interna de las escamas						3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
216			Semillas aladas			9			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
217	c		Lóngitud de las semillas (mm)																							
218				Max													50	7	5	50	8	5	8	8	8	
219				Min													3	6	4	15	5	3	3	7	3	
220			Ala de la semilla			9																				
221			bien desarrollada			55			1	0	0	0	0	0	0	0	1	3	3	-1	0	3	3	3	3	
222			muy estrecha o rudimentaria			55			1	0	0	0	0	0	0	0	1	-1	-1	3	0	-1	-1	-1	-1	
223			caduca			56			1	0	0	0	0	0	0	0	1	-1	-1	3	0	-1	-1	-1	-1	
224			persistente			56			1	0	0	0	0	0	0	0	1	3	3	-1	0	3	3	3	3	
225	c		Longitud del ala de la semilla (mm)																							
226				Max														28	14			17	20	30	20	
227				Min														22	10			12				

Apéndice 2. Tabla Excel de conocimiento de determinación taxonómica.

Orden	"c" cuantitativa, " " cualitativa	Agrupación	Características	Tipo dato cuantitativo	Agrupaciones ocultas	Grupo auto-excluyentes	Aproximadas	Orden relación proximadas	PINACEAE	Abies	Abies pinsapo	Abies alba	Pseudotsuga	Pseudotsuga menziesii	Picea	Picea abies	Pinus	Pinus halepensis	Pinus uncinata	Pinus pinea	Pinus radiata	Pinus sylvestris	Pinus nigra	Pinus pinaster	Pinus nigra ssp. nigra	Pinus nigra ssp. salzmannii
228		Hábitat																								
229			Suelos silíceos o descalcificados			57			1	0	0	0	0	0	0	0	1	1	1	0	3	1	1	3	0	0
230			Sustrato calcareo			57			1	0	0	0	0	0	0	0	1	2	1	0	-1	1	2	0	0	0
231			Ambiente costero o seco y soleado			58			1	-1	-1	-1	0	0	0	0	1	3	-1	0	0	0	0	0	0	0
232			Zonas altas de montaña			58			1	3	3	3	0	0	0	0	1	-1	3	0	0	0	0	0	0	0
233		Genética																								
234			Número de cromosomas (2n)																							
235				Max							24	26	26	24	24	24		24		24	24	24		24		
236				Min							24	26	26	22	22			24		24	24	24		24		

Apéndice 3. Descripción de esquemas XML. Representación del conocimiento

El experto botánico que realiza la introducción de datos al conocimiento que utilizará el sistema de determinación taxonómica, utiliza una tabla Excel para dicho fin. De esta manera proporcionamos al usuario una herramienta conocida con la cual se siente cómodo para trabajar. Pese a esta situación, el sistema no utiliza directamente el Excel para la operativa que realiza. El sistema transforma la información obtenida de las tablas y la representa mediante XML. De esta manera la información queda almacenada en un repositorio de XMLs. El objetivo es tener una representación final de la información bajo este estándar de intercambio estructurado de datos, que nos permitirá mayor flexibilidad a la hora de evolucionar el sistema, comunicarle con otros sistemas y a fin de divulgar el conocimiento

En este apéndice presentamos la definición de los dos complejos esquemas XML diseñados para recoger de forma estructurada la información contenida en las tablas.

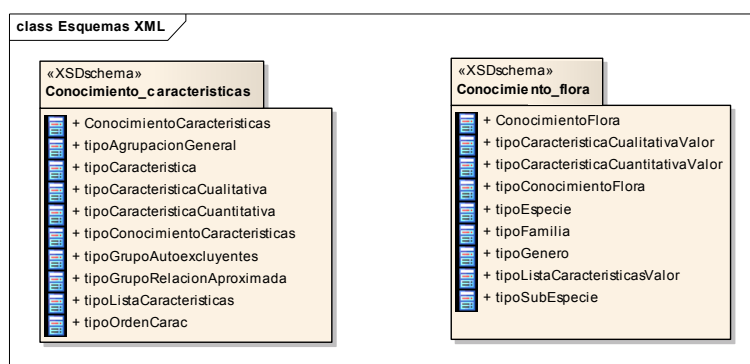


Figura 12-1. Esquemas XML

Conocimiento características

Esquema que recoge la información de las características, se ha estructurado de tal manera que contemple todas las casuísticas derivadas de su representación en la tabla Excel.

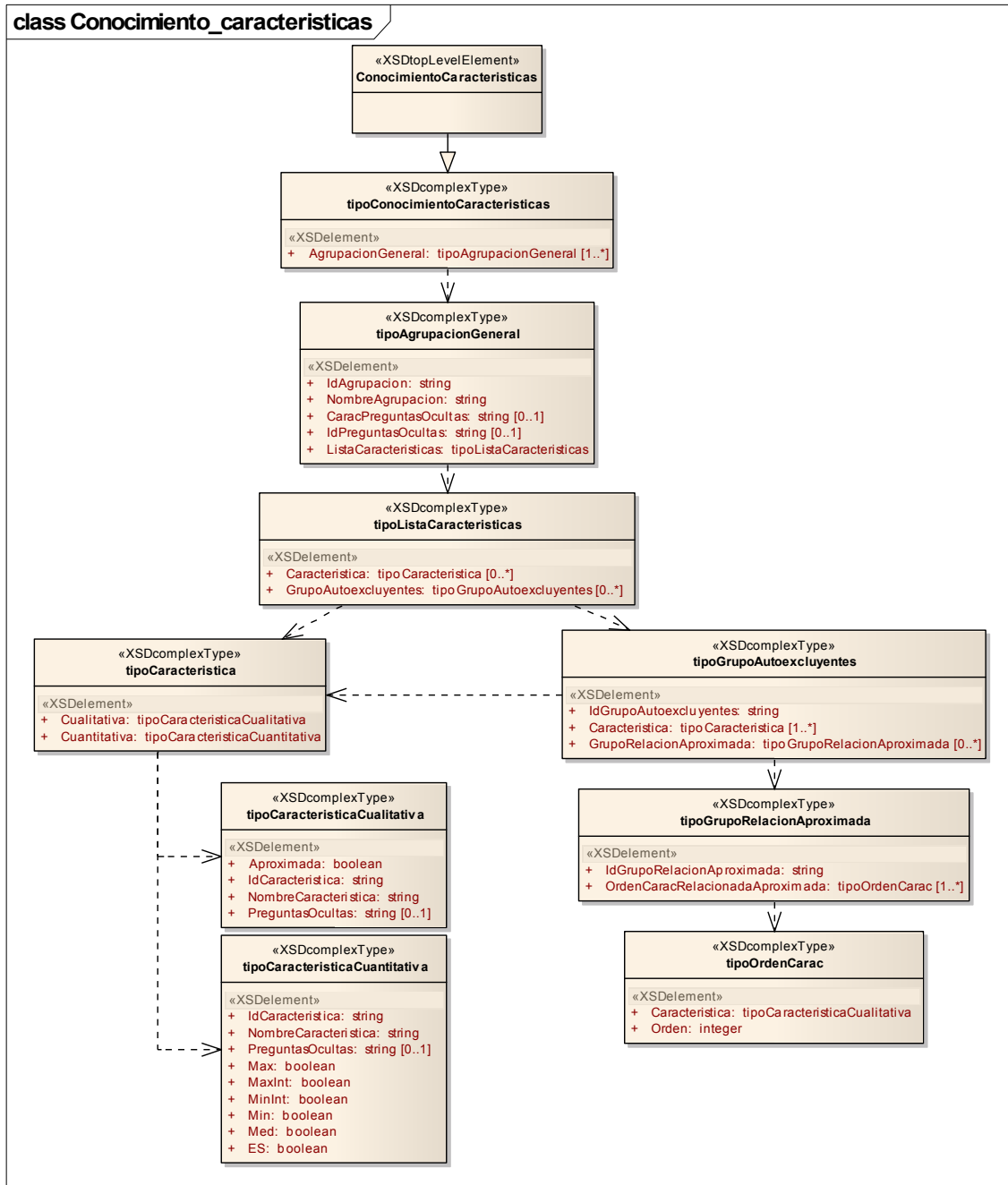


Figura 12-2. Conocimiento_caracteristicas

ConocimientoCaracteristicas

Conocimiento de las posibles características y su agrupación.

- **tipoAgrupacionGeneral**

Agrupación general de características.

Atributo	Descripción
Nombre: IdAgrupacion Tipo: string	Identificador único de la agrupación
Nombre: NombreAgrupacion Tipo: string	Nombre que identifica a la agrupación. Por ejemplo Biotipo, Copa, Sabia...
Nombre: CaracPreguntasOcultas Tipo: string [0..1]	Nombre de la característica que la hace permanecer oculta.
Nombre: IdPreguntasOcultas Tipo: string [0..1]	Identificador de preguntas ocultas
Nombre: ListaCaracteristicas Tipo: tipoListaCaracteristicas	Lista de características de la agrupación.

- **tipoCaracteristica**

El tipo característica engloba las dos posibilidades que tenemos, características cuantitativas o cualitativas.

Atributo	Descripción
----------	-------------

Atributo	Descripción
Nombre: Cualitativa Tipo: tipoCaracteristicaCualitativa	Característica cualitativa.
Nombre: Cuantitativa Tipo: tipoCaracteristicaCuantitativa	Característica cuantitativa.

- **tipoCaracteristicaCualitativa**

Información asociada a cada una de las características cualitativas.

Atributo	Descripción
Nombre: Aproximada Tipo: boolean	Indica que la característica cualitativa es aproximada, se puede definir por los valores poco, medio, mucho.
Nombre: IdCaracteristica Tipo: string	Identificador único de la característica.
Nombre: NombreCaracteristica Tipo: string	Nombre de la característica.
Nombre: PreguntasOcultas Tipo: string [0..1]	Número asignado en el excel para esta característica. Indica que tiene preguntas que dependen de ella.

- **tipoCaracteristicaCuantitativa**

Información asociada a cada una de las características cuantitativas.

Atributo	Descripción
Nombre: IdCaracteristica Tipo: string	Identificador único de la característica.
Nombre: NombreCaracteristica Tipo: string	Nombre de la característica.
Nombre: PreguntasOcultas Tipo: string [0..1]	Número asignado en el excel para esta característica. Indica que tiene preguntas que dependen de ella.
Nombre: Max Tipo: boolean	Máximo del intervalo. Indica si la característica tiene (true) o no tiene (false) "Max" como dato a rellenar por el usuario.
Nombre: MaxInt Tipo: boolean	Indica si la característica tiene (true) o no tiene (false) "MaxInt" como dato a rellenar por el usuario.
Nombre: MinInt Tipo: boolean	Indica si la característica tiene (true) o no tiene (false) "MinInt" como dato a rellenar por el usuario.
Nombre: Min Tipo: boolean	Mínimo del intervalo. Indica si la característica tiene (true) o no tiene (false) "Min" como dato a rellenar por el usuario.
Nombre: Med Tipo: boolean	Media

Atributo	Descripción
Nombre: ES Tipo: boolean	Error standard

- **tipoConocimientoCaracterísticas**

Conocimiento de las posibles características.

Atributo	Descripción
Nombre: AgrupacionGeneral Tipo: tipoAgrupacionGeneral [1..*]	Agrupación de características.

- **tipoGrupoAutoexcluyentes**

Grupo de características auto-excluyentes.

Atributo	Descripción
Nombre: IdGrupoAutoexcluyentes Tipo: string	Identificador del grupo de Auto-excluyentes
Nombre: Caracteristica Tipo: tipoCaracteristica [1..*]	Características incompatibles.
Nombre: GrupoRelacionAproximada Tipo: tipoGrupoRelacionAproxim	Grupos de relación aproximada de características.

Atributo	Descripción
ada [0..*]	

- **tipoGrupoRelacionAproximada**

Grupo de características relacionadas aproximadas.

Atributo	Descripción
Nombre: IdGrupoRelacionAproximada Tipo: string	Identificador del grupo de relación aproximada.
Nombre: OrdenCaracRelacionadaAproximada Tipo: tipoOrdenCarac [1..*]	Característica cualitativa relacionada aproximada junto con su orden de aparición en la escala.

- **tipoListaCaracteristicas**

Lista de características.

Atributo	Descripción
Nombre: Caracteristica Tipo: tipoCaracteristica [0..*]	Podrá tener de 0 a infinito características.
Nombre: GrupoAutoexcluyentes Tipo: tipoGrupoAutoexcluyentes	Lista de grupos de características Autoexcluyentes entre si.

Atributo	Descripción
[0..*]	

- **tipoOrdenCarac**

Pares de característica y orden en la relación aproximada. Es decir, el orden en el que aparecerá sobre la escala.

Atributo	Descripción
Nombre: Característica Tipo: tipoCaracteristicaCualitativa	Característica del grupo de relación aproximada.
Nombre: Orden Tipo: integer	Orden de la característica sobre la escala del grupo de relación aproximada.

Conocimiento flora

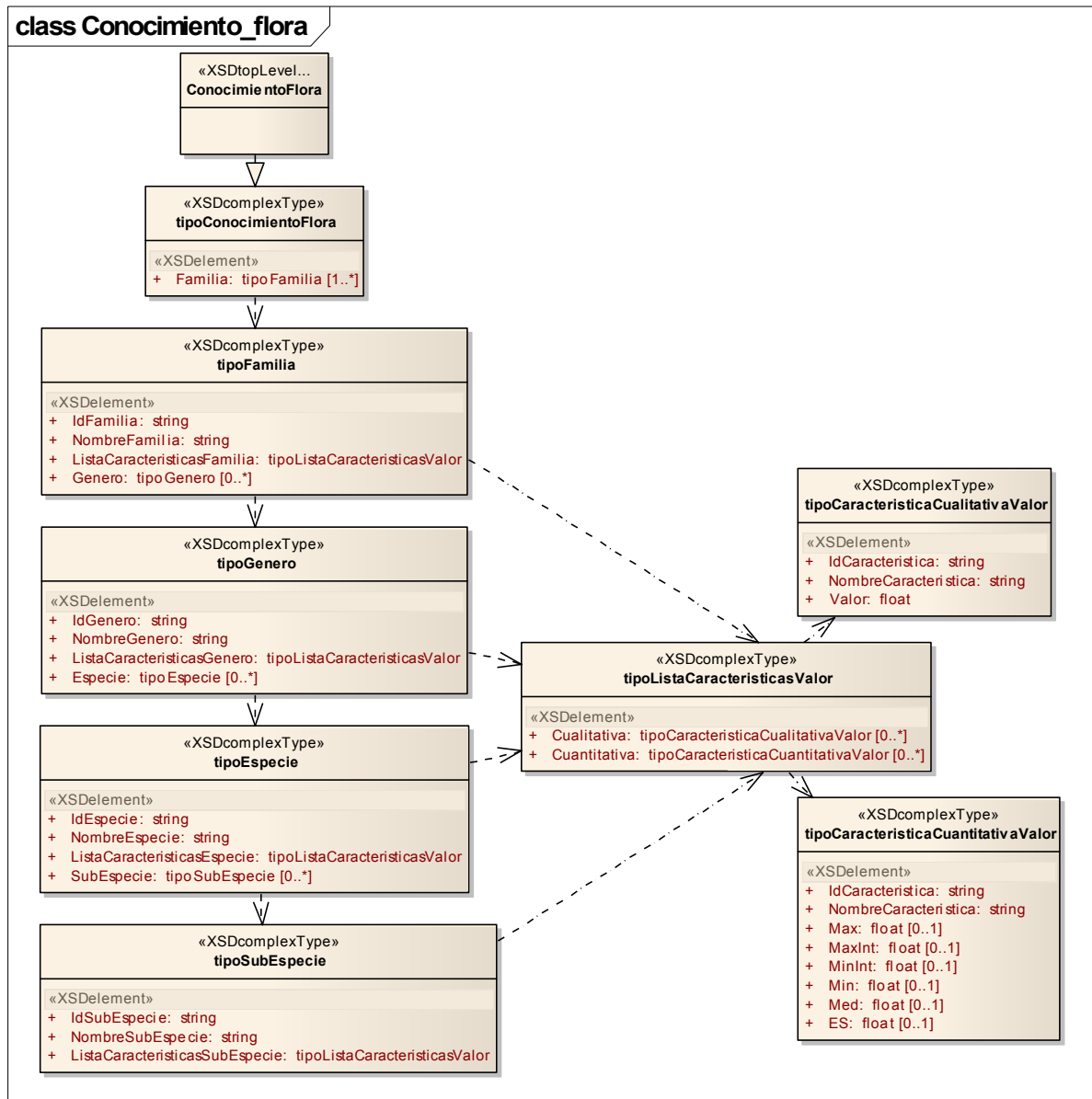


Figura 12-3. Conocimiento_flora

ConocimientoFlora

Conocimiento de las familias, géneros y especies.

- **tipoCaracteristicaCualitativaValor**

Característica cualitativa y su valor asignado.

Atributo	Descripción
Nombre: IdCaracteristica Tipo: string	Identificador de la característica.
Nombre: NombreCaracteristica Tipo: string	Nombre de la característica.
Nombre: Valor Tipo: float	Valor de la característica.

- **tipoCaracteristicaCuantitativaValor**

Característica cuantitativa y su valor asignado.

Atributo	Descripción
Nombre: IdCaracteristica Tipo: string	Identificador de la característica.
Nombre: NombreCaracteristica Tipo: string	Nombre de la característica.
Nombre: Max Tipo: float [0..1]	Máximo

Atributo	Descripción
Nombre: MaxInt Tipo: float [0..1]	Máximo del intervalo.
Nombre: MinInt Tipo: float [0..1]	Mínimo del intervalo.
Nombre: Min Tipo: float [0..1]	Mínimo
Nombre: Med Tipo: float [0..1]	Media
Nombre: ES Tipo: float [0..1]	Error estandard

- **tipoConocimientoFlora**

Conocimiento de las familias, géneros y especies.

Atributo	Descripción
Nombre: Familia Tipo: tipoFamilia [1..*]	Familias.

- **tipoEspecie**

Cada una de las species.

Atributo	Descripción
Nombre: IdEspecie Tipo: string	Identificador único de la especie.
Nombre: NombreEspecie Tipo: string	Nombre de la especie
Nombre: ListaCaracteristicasEspecie Tipo: tipoListaCaracteristicasValor	Lista de características de la especie con su valor.
Nombre: SubEspecie Tipo: tipoSubEspecie [0..*]	Lista de subespecies asociadas a la especie.

- **tipoFamilia**

Cada una de las familias.

Atributo	Descripción
Nombre: IdFamilia Tipo: string	Identificador único de la familia.
Nombre: NombreFamilia Tipo: string	Nombre que identifica a la familia.
Nombre: ListaCaracteristicasFamilia Tipo: tipoListaCaracteristicasValor	Lista de características de la familia.

Atributo	Descripción
Nombre: Genero Tipo: tipoGenero [0..*]	Lista de géneros de la familia.

- **tipoGenero**

Cada uno de los generous.

Atributo	Descripción
Nombre: IdGenero Tipo: string	Identificador único del género
Nombre: NombreGenero Tipo: string	Nombre que identifica al género.
Nombre: ListaCaracteristicasGenero Tipo: tipoListaCaracteristicasValor	Lista de características del género.
Nombre: Especie Tipo: tipoEspecie [0..*]	Especies asociadas al género.

- **tipoListaCaracteristicasValor**

Asociación de cada característica con su valor.

Atributo	Descripción
Nombre: Cualitativa	Asociación de cada característica cualitativa con su valor.

Atributo	Descripción
Tipo: tipoCaracteristicaCualitativa Valor [0..*]	
Nombre: Cuantitativa Tipo: tipoCaracteristicaCuantitativa Valor [0..*]	Características cuantitativas.

- **tipoSubEspecie**

Informació de cada Subespecie.

Atributo	Descripción
Nombre: IdSubEspecie Tipo: string	Identificador único de la SubEspecie
Nombre: NombreSubEspecie Tipo: string	Nombre de la SubEspecie
Nombre: ListaCaracteristicasSubEspecie Tipo: tipoListaCaracteristicasValor	Lista de características de la especie con su valor.